Fluorosequencing implications for COVID-19 and viral outbreaks

Jagannath Swaminathan, Cofounder of Erisyon Inc. – jagannath@erisyon.com

- **Erisyon is commercializing the world's first single molecule protein sequencer. The technology brings the advantages of next generation genomic sequencing to proteomics including high sensitivity, massively parallel throughput, and digital quantification.**
- **Pandemic preparedness and response is a key part of protecting the world from another outbreak like the current COVID-19 crisis. Techniques to monitor and intervene in zoonotic viral outbreaks before they make their way from non-human animals to humans are critical first line defenses that should be put in place. In the case a pandemic does break out, the ability to quickly characterize the virus's infectious vector is essential.**
- **Erisyon's technology, known as fluorosequencing, can contribute to these efforts. Fluorosequencing's ability to analyze glycosylation proteins like COVID-19's spike protein is a powerful tool for the development of tests, treatments, and vaccines.**

Background - Viral Glycosylation

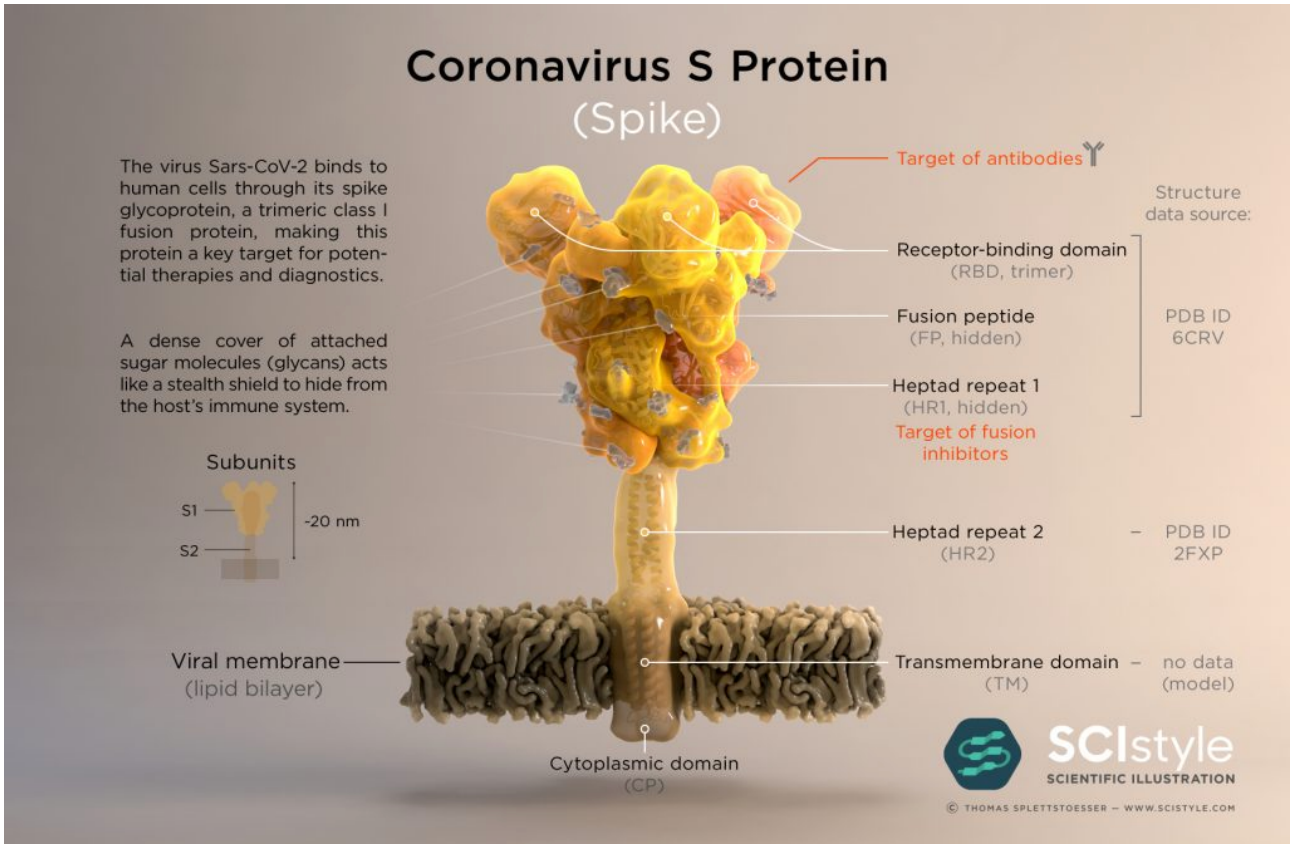
Viruses are nanometer- sized infectious agents that replicate by invading a living cell and using its machinery to reproduce. Ranging from 20nm to up to 500nm in size, it is estimated that there are around 1x1031 total viruses on Earth, or about 10 billion times more stars than in the known universe. If all of the world's viruses were lined up end to end, they would stretch for over 100 million light years.

Every species on Earth - flora and fauna - is susceptible to a viral infection. It's unknown how many species of viruses exist, but estimates range from the millions to even the billions. Today, we have identified 214 species that are known to infect humans. Of these species, only 34 are zoonotic, meaning they have the ability to infect non-human hosts. These zoonotic species can incubate and evolve in animals like birds, rodents, bats, or pigs before they make the leap to humans. In such instances, pandemics become possible as the virus's novelty can overcome or evade the natural defenses of the human immune system.

SARS-Cov-2 (aka COVID-19), a member of the coronavirus species and, the virus that is currently paralyzing the world, is precisely such a zoonotic virus. Currently suspected to have originated in bats and passed to humans through pangolins, COVID-19 has been declared a pandemic by the World Health Organization (WHO). As of March 30 2019, there are over 775,000 confirmed cases globally of which over 33,000 have been fatal.

How a dangerous zoonotic virus like COVID-19 is able to avoid the human immune system is a subject of intense study. Every biological system is made of a complex and chaotic set of variables that are interrelated, nearly impossible to differentiate, and include environmental, biological and societal factors.

One target of intense scrutiny is the so-called spike protein that is expressed on the outer envelope of COVID-19. This molecule protrudes from the outer membrane like the points on a crown (hence the origin of "corona" in the name, which is Latin for crown.) The spike protein (also known as the S protein) has two ends - one binds to the human cell while the other end pulls the virus closer to the cell. This protein is decorated by a number of complex sugar chains that attach to specific locations on the molecule by a process called glycosylation. The precise positions where these sugars are attached to the spike protein are suspected to simultaneously allow the COVID-19 to evade the immune system while conferring the ability to attach itself to a specific location on a cell in order for the virus to infect itself.



Glycosylation is what is known as a post-translational modification (PTM). PTMs provide crucial functionality for proteins including energy to catalyze a reaction, support to change shape or, in the case of COVID-19, provide functionality to provide access to a site like a key might for a lock. By definition, PTMs take place after a protein is translated from RNA which had previously been transcribed from DNA. As a result, since a genomic sequence is an instruction set for how to assemble a protein, it provides no information on what happens after the protein has been made.

However, different kinds of PTMs only modify certain amino acids, which are the constituent parts of a protein. For instance, a glycosylation event can add a sugar molecule to the amino acid Asparagine while another type of PTM, phosphorylation, can only modify serines, threonines, and

tyrosines. So what the genomic sequence does inform us is how many potential PTMs there may be, and where they might be located.

What is subsequently needed is a technique for understanding the precise number and locations of PTMs in order to understand how many, where, and what type of sugars have been added to the protein. Such a technique would have broad and immediate impact in all areas of life science research including oncology, plant science and, in the case of our current crisis, pandemic preparedness and response.

Once the genomic sequence of COVID-19 was completed, the map for where sugar molecules might be located on the spike protein would need to be analyzed. With such information, researchers and clinicians might be able to design vaccines or prescribe medications that target and deactivate COVID-19.

Unfortunately, the existing proteomic techniques for analyzing glycosylation are either inadequate, difficult to use, or provide data that is hard to interpret. The most common approach is to use what are known as affinity assays. Typically made of antibodies, affinity assays should be able to bind to locations on a target with high specificity and precision. Practically, affinity assays produce qualitative results depending on the quality of the affinity agent, the target location, the testing conditions, and a host of other considerations. Furthermore, an affinity assay can only identify a single target at a time and cannot provide a holistic perspective of the PTMs on a protein.

The other technique available is mass spectrometry. A marvel of physics, a mass spectrometer provides exquisitely precise information about a molecule by measuring its mass-to-charge ratio. From this information, it's possible to calculate the characteristics of a sample and identify its constituents. Unfortunately, mass spectrometry is poorly suited for quantitative analysis. Mass-spec identifies protein fragments by weighing the fragments with high precision (technically it measures the mass-to-charge-ratio, but it can be imagined as a weight). Mass-spec then compares the weight of the found molecule to a database of possible weights to make its identification. Unfortunately, glycosylation is often characterized by random assortments of sugars and therefore the mass-spec data produces probabilistic models comprising possible permutations of sugar branches, and is often not quantitative (i.e it is difficult to estimate how many of those proteins had the glycosylation).

Erisyon's single molecule protein sequencing technology, known as fluorosequencing, provides the level of sensitivity, resolution, and specificity necessary for this kind of analysis. Emerging from the Marcotte and Anslyn labs at UT Austin, Erisyon is bringing many of the benefits of next generation genomic sequencing to proteomics including high sensitivity, massively parallel throughput, and digital quantification.

Erisyon's Technology

Fluorosequencing works by combining a number of proven and scalable techniques including classical Edman degradation, single molecule microscopy, and advanced machine learning. Described as "a marriage across the ages" fluorosequencing combines techniques from the beginning of the age of biotechnology with today's most cutting edge analytical research in a way that is more than 1,000,000 times more sensitive than the state of the art, mass spectrometry.

Erisyon describes fluorosequencing operative strategy as the "protein Wheel of Fortune." In the game show The Wheel of Fortune, a contestant must answer an incomplete puzzle that has just a few letters. Their only guidance is a knowledge of the alphabet, the English dictionary, and the category of the phrase. This information is usually enough to make a highly accurate educated guess.

Fluorosequencing's approach is to identify all of the locations of just a few amino acids or PTMs on a peptide and, in combination with a knowledge of the possible proteins in a sample, make a highly accurate, educated call on the protein that is being sequenced. This process was detailed in a paper in the very prestigious journal Nature Biotechnology in December 2018.

The first step in fluorosequencing is preparing a sample with millions or even billions of peptides (fragments of proteins) by labeling all of a select number of amino acids or PTMs, usually just three or four, with fluorescent dyes. The labeled peptides are then immobilized on a glass slide by their C-terminal ("bottom") while the N-terminal ("top") remains exposed. The slide is then excited by three or four channels of laser light and an image is taken through a microscope. The image appears like a sky full of constellations on a dark night, with each "star" an individual peptide. The intensity of light in each of the three or four laser channels is measured and recorded.

The next step is to perform an Edman degradation cycle. Edman degradation is a chemical process invented by Pehr Edman, a Swedish scientist, in 1950. For decades it was the gold standard for protein analysis until more modern techniques like mass spectrometry were developed. It works by removing the top amino acid, and only the top amino acid, during each cycle. Once the Edman cycle has finished, another image is taken to measure the intensity of fluorescence in each channel once again.

If the intensity signatures in one of the channels havehas changed, then it can be confirmed that the amino acid or PTM which was labeled with that dye was in that position. If nothing has changed, that spot is registered with an unknown. Then another Edman cycle is executed, followed by imaging and intensity measurement, and so on.

When the peptide has been exhausted, the resulting incomplete sequence, or fluorosequence, is matched against a database to identify which protein it came from. Since millions or even billions of peptides were sequenced in parallel, a comprehensive sense of all of the proteins in the sample can be achieved.

Pandemic Preparedness and Response

The applications of Erisyon's technology are wide and varied. It can be applied including diagnostics for neurological disorders, identifying targets for immuno-oncology therapy, isolating low level biomarkers for liquid biopsies, and dozens of other use cases. In the case of pandemic preparedness and response there applications immediately reveal themselves:

Pandemic Preparedness

As previously described, a virus's mutation may add, delete, or reposition asparagine amino acids, which in turn can alter the glycosylation positions and patterns of a virus. As part of pandemic preparedness, the glyco-signature of zoonotic viruses collected from non-human animal species can be monitored and catalogued prior to their interspecies leap. These signatures can be compared to or assayed against human models to predict how infectious they may or may not be. If a virus is deemed to have a human vector, then tests, treatments, and vaccines can be developed well before an outbreak occurs.

Pandemic Response

In case an outbreak does take place, it is paramount to quickly gather the most accurate and complete information. Fluorosequencing can be used to quickly analyze the glycosylation patterns of a virus's infectious vector. Such information can assist in producing tests, treatments, and vaccines.

Pandemic Tracking

A virus can continue to mutate even during a pandemic andor can affect certain parts of the population in different ways. While the data is incomplete, variable, and at times even anecdotal, COVID-19 has certainly affected certain parts of the world differently. There is even some evidence that people with certain blood types are more susceptible to severe infection than others. Blood types are defined by the types and positions of glycosylation on the red blood cell. Associations such as these, and other emerging possibilities, can be analyzed more completely using fluorosequencing.

References

1. Microbiology by numbers. Nat Rev Microbiol 9, 628 (2011). https://doi.org/10.1038/nrmicro2644
2. Woolhouse, M., Brierley, L. Epidemiological characteristics of human-infective RNA viruses. Sci Data 5, 180017 (2018). https://doi.org/10.1038/sdata.2018.17
3. Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. The proximal origin of SARS-CoV-2. Nat Med (2020). https://doi.org/10.1038/s41591-020-0820-9
4. https://www.who.int/emergencies/diseases/novel-coronavirus-2019
5. https://www.bing.com/covid
6. David J. Vigerust, Virginia L. Shepherd, Virus glycosylation: role in virulence and immune interactions, Trends in Microbiology, Volume 15, Issue 5, 2007, Pages 211-218, ISSN 0966-842X, https://doi.org/10.1016/j.tim.2007.03.003.
7. Ieva Bagdonaite, Hans H Wandall, Global aspects of viral glycosylation, Glycobiology, Volume 28, Issue 7, July 2018, Pages 443–467, https://doi.org/10.1093/glycob/cwy021
8. Swaminathan J, Boulgakov AA, Hernandez ET, et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures [published online ahead of print, 2018 Oct 22]. Nat Biotechnol. 2018;10.1038/nbt.4278. doi:10.1038/nbt.4278
9. Collins, B., Aebersold, R. Proteomics goes parallel. Nat Biotechnol 36, 1051–1053 (2018). https://doi.org/10.1038/nbt.4288
10. Swaminathan J, Boulgakov AA, Hernandez ET, et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures [published online ahead of print, 2018 Oct 22]. Nat Biotechnol. 2018;10.1038/nbt.4278. doi:10.1038/nbt.4278
11. Relationship between the ABO Blood Group and the COVID-19 Susceptibility Jiao Zhao, Yan Yang, Hanping Huang, Dong Li, Dongfeng Gu, Xiangfeng Lu, Zheng Zhang, Lei Liu, Ting Liu, Yukun Liu, Yunjiao He, Bin Sun, Meilan Wei, Guangyu Yang, Xinghuan Wang, Li Zhang, Xiaoyang Zhou, Mingzhao Xing, Peng George Wang medRxiv 2020.03.11.20031096; doi: https://doi.org/10.1101/2020.03.11.20031096