Medical Case Reports and Reviews

Short Communication ISSN: 2517-7214

A strategy for rapidly making a vaccine and treatment for the disease caused by the Wuhan-Corona Virus 2019 (COVID19)-Part two

George Pieczenik*

Rutgers University, School of Environmental and Biological Sciences, New Brunswick, N.J. 08901, USA

Communication

Rutgers University, School of Environmental and Biological Sciences, New Brunswick, N.J, 08901, USA

Our previous paper Part 1 [1] identified 184 possible epitopes for COVID19 based on amino acid sequence homology to SARS's epitopes.

In this part, we will use a more sophisticated analysis which we call the Kimura Peptide Analysis of Synonymous Codon Usage (KPASCU). This analysis was developed with Sydney Brenner as part of our previous work with Francis Crick and Sir Aaron Klug [2].

Presently, a majority of bioinformaticians analyze sequences by homology between the amino acid sequences of proteins. The assumption in this type of analysis is that if an identical nucleotide sequence codes for an identical peptide sequence in two different proteins in two different viruses; then, this is evidence of "descent from a common ancestor". This is the Darwinian assumption and is the most common method of analysis. Just recently, for example, the following sequence VLLFLAFVV was identified as a common epitope in the Envelope protein (E) of both SARS and COVID19 [3] and they suggest this as the epitope for the basis of a vaccine. Their work is quite pretty. However, we feel that this may not be the best epitope on which to base a vaccine. We have two reasons. Firstly, this sequence is completely hydrophobic and that is rare in antibody-epitope interactions.

The second reason is that "descent from a common ancestor" is not the right interpretation of sequences to identify an epitopic region that can be the basis of a vaccine. The sequence they propose was identified in our previous paper Part 1 as one of the 184 common epitopes. In fact, an examination of all the 184 possible epitopic regions from our previous paper [1] gives a larger exposed surface region in this E protein as LIVNSVLLFLAFVVFLLVTLAILTALRLCAY. This means that their epitopic sequence 9 amino acids long is part of a larger sequence 31 amino acids long that is on the surface of the Envelope protein and accessible to antibodies.

Figure 1 shows the relationship between the nucleotide sequences coding for both SARS and COVID19 in this region. See below.

The first thing that is noticeable is that the nucleotide codings for this region in both SARS and COVID19 are identical except for one synonymous nucleotide change at the fourth Valine. In SARS, this Valine is coded by GUC while in COVID19 this Valine is coded by GUU (at the RNA level, Fig. 1 gives the DNA sequence). This is a simple transitional mutation of a C to U between the two viruses. Interestingly, a single C to U change in the polio virus is a change from the Leon strain (pathogenic) to the Sabin vaccine (non pathogenic) [4]. Therefore, a single change in the right place can convert a pathogenic virus into a non pathogenic virus i.e. a live virus vaccine (Sabin). However, this change between SARS and COVID19 did not covert the COVID19 into a non pathogenic strain [*].

The almost perfect coding identity shown in Figure 1 between these two regions of SARS and COViD19 suggests that these regions evolved as "descent from a common ancestor". Therefore, one can conclude that there was no particular coding selection imposed on the nucleotide sequence, nor any binding selection on the coded surface amino acids.

While their work [3] is bioinformationally exceptionally pretty, we feel that this is not a strong candidate for a vaccine as there is no evidence for the amino acid sequence being selected for a function that is independent of its being coded. It is just "descent from a common ancestor".

Evidence for which COVID19's surface amino acids bind strongly.

What would evidence be for an amino acid sequence being selected for having a unique function, such as, a strong binding function, that is independent of its coding sequence?

One type of evidence is "convergent evolution" of amino acid sequences rather than "descent from a common ancestor". For example, tapirs and pigs look alike but tapirs are odd-toed and pigs are even-toed ungulates [5]. At the molecular level, "convergent evolution" would look like identical amino acid sequences among the epitopes of SARS and COVID19, but the nucleotide sequences coding for any two identical epitopes would be not be identical. Their codings would be quite different.

In other words, different synonymous codons coding for the same amino acid sequence would be evidence of "convergent evolution". The convergence of amino acid sequences is evidence that selection is on the amino acid and not on the nucleotide sequence. It would also be evidence that the nucleotide sequences are not related and not descended from a common ancestor.

Correspondence to: George Pieczenik, Rutgers University, School of Environmental and Biological Sciences, New Brunswick, N.J. 08901, USA, E-mail: piecze@sebs.rutgers.edu

Received: February 24, 2020; **Accepted:** March 02, 2020; **Published:** March 05, 2020

Experimental Evidence for Functional Convergent Evolution-two different nucleotide sequences evolved coding for similar amino acid sequence which were selected for the function of binding to an identical target.

We ran an experiment where we actually generated a universe of different peptides on the surface of filamentous phage ranging in size from 7 to 12 [6,7] and bound them to a specific defined 7 amino acid target (the TNF alpha human epitope). Then, the bound phage were purified, amplified and rebound several times. The results are shown below in Figure 2.

Two phages given above were identified which coded for 4 identical amino acid sequences in exactly the same location in respect to each other.

The four amino acid sequences are H R L X D. (His-Arg-Leu-N-Asp) One phage binding sequence was identified in the 7 amino acid combinatorial library, the second phage binding sequence was identified in the 12 amino acid binding library. The statistical significance of independently getting the same 4 amino acids in the same location from both differently sized libraries is less than 1 in 160,000.

The results show that the same amino acid sequence converged as a binding sequence to the target epitope, while the sequence coding for them were different. That is different codons coded for the same sequence. This is evidence that selection was for the amino acid sequence and not for the nucleotide sequence. We shall call these Kimura peptides, because Kimura proposed the fixation of synonymous codings as random [8] This assumption is used to calculate mutation rates.

This is the first and only direct experimental test of Kimura's hypothesis of neutral mutations. All other evidence for Kimura's hypothesis has been *a posterior*i biometric analysis.

What we have presented here is a functional direct test that selection can occur at the protein level with neutral fixation of synonymous coding. So at least Kimura is right on the possibility that his and Jukes [9] neutral mutation model of evolution can be part of the evolution of protein sequences at least in terms of their binding functions.

Footnote: This binding sequence HRLxD is a subset of a sequence which binds the human TNF α epitope. This full sequence can replace Humira without the lymphoma side effects of Humira.

The Novel Use of the Kimura Peptide Analysis Method

The use of the Kimura Peptide Analysis method for defining effective epitopes in unknown viruses is novel.

If the same sequence is coded by different synonymous codings in related viruses, then, it means that selection is on the protein or peptide region. In specific, then if the region is known to be a surface peptide region, such as an epitope or a series of collinear epitopes, then it argues very strongly that the region is being selected by a binding interaction with some necessary receptor or binding dock for the adaptive survival of the virus. However, on the other hand, if it was being selected by host antibodies; then, the virus would not survive. Without viral survival there would be no disease. Because it was originally selected as a surface peptide for some important survival function of the virus such as receptor binding or membrane insertion, then, it later became a target for host antibodies.

Therefore, we predict the following epitopes as necessary to the survival of COVID19 and the proper epitopes from which to make a vaccine and/or a neutralizing binding peptide. We can do this by analyzing the known 184 epitopes in common with SARS and COVID19.

Very importantly, this epitopic sequences should be confirmed by mapping the epitope binding of antibodies from patients who recovered from infection by COVID10 infection (Figure 3). New England Biolabs has published protocols for identifying the binding sequence of monoclonals by screening them with phage display libraries [7].

8/16 50 percent and one diff aa.

Synonymous Differences

Envelope 1/31 3%

Membrane 4/10 40%

SPIKE 12/23 52%

SPIKE S2 8/13 61%

The Spike Region S2 (RBD) was predicted to be a binding region by microscopic visualization using cryo-EM. (However, this region in COVID 19 differs by one amino acid from SARS in its epitopic function even though it is coded by 8 different synonymous codons. Three monoclonals that bound SARS did not bind the homologous COVID19 region.

tta ata gtt aat agc gta ctt ctt ttt ctt gct ttc gtg gta ttc ttg cta **gtc** aca cta gcc atc ctt act gcg ctt cga ttg tgt gcg tac SAR

L I V N S V L L ***F L A F V V F L L*** **V** **T L** A I L T A L R L C A Y

tta ata gtt aat agc gta ctt ctt ttt ctt gct ttc gtg gta ttc ttg cta **gtt** aca cta gcc atc ctt act gcg ctt cga ttg tgt gc gcg tac CV19

Figure 1. Envelope Protein (E) comparison of SARS and COVID19

12 amino acid display phage binding TNF α epitope (Humira's target)

GCT GGT GTG TAT CCT **CAT CGT TTG** TCT **GAT** GTG CTT

A G V Y P **H R L** S **D** V L

GCG ATG **CAT CGG CTT** GGG **GAT**

A M **H R L** G **D**

7 amino acid display phage binding TNF α epitope (Humira's target)

Figure 2. Results of the binding TNF α epitope (Humira's target) for two phages

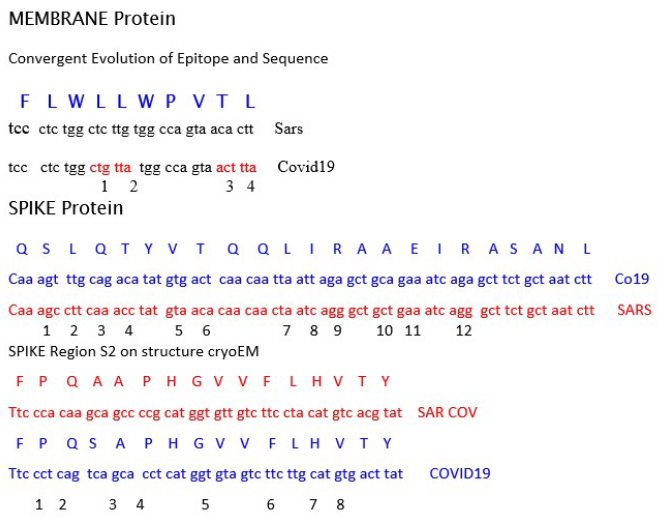

Figure 3. Summary of three epitopic sequences by Kimura Peptide analysis

Clearly, the one amino acid difference is significant and rules out this as an epitope for COVID19 even though there alot of different codon usuage.

Spike Two- This region is 26 amino acids from Spike S2 and is the most likely epitopic candidate for developing a vaccine.

Because it is over 12 amino acids, it is also immunogenic and can be used as a peptide stimulant to generate and antibody against both it and COVID19. Therefore, this peptide or subsets of this peptide can be used to make a vaccine and/or directly stimulate and antibody response (Q S L Q T Y V T Q Q L I R A A E I R A S A N L).

In addition, the following sequences and subsets can be used as targets for binding ligands made from a combinatorial library [6]. These ligands can then replace monoclonals if used as passive immunization.

1) F L W L L W P V T L

2) Q S L Q T Y V T Q Q L I R A A E I R A S A N L

3) F P Q S A P H G V V F L H V T Y

These ligands can be made much quicker than monoclonals, which should also be made. These ligands can then neutralize the virus and its ability to infect.

In addition, the DNA sequence of these codings can be inserted into adenovirus vectors in order to create a competitive virus creating antibodies to the COVID19 virus.

Again, it should be re-iterated, that the epitopic mapping of antibodies derived from infected patients using combinatorial libraries will identify epitopes that can be used to make a vaccine. If the epitopes identified by combinatorial mapping corresponds to the epitopes identified by Kimura Peptide Analysis, then one can be assured that the right neutralizing epitopes have been identified.

In addition, it should be noted that the binding epitope on COVID19 is "stronger" than the binding epitope on SARS. There are more synonymous codings. This means the virus while not as virulent as SARS would be more rapidly infectious. The virulence diminution may be explained by the synonymous C to U change in the E protein, which is similar to the diminution in the Leon strain of polio into the Sabin strain at position 472. This is a GC to GU base pair change in polio and will probably be similar in the SARS to COVID19 change.

*As an alternate strategy, isolating a competitive non pathogenic strain of COVID19 will be possible as the epidemic attenuates. We have previously isolated a non pathogenic competitive strain of HIV in addressing the AIDS epidemic. (Scolaro, M., Durham, R. and Pieczenik, G. (1991) "Potential Molecular Competitor for HIV", The Lancet, Vol.337, p.731).

References

1. Pieczenik G, Qi T, Zylstra G (2020) "A strategy for rapidly making a vaccine and treatment for the disease caused by the Wuhan-Corona Virus (WCV)" *Med Case Rep Rev* 3: 1-1.
2. Crick FH, Brenner S, Klug A, Pieczenik G (1976) A speculation on the origin of protein synthesis. *Origins of life* 7: 389-397.
3. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, et al. (2020) Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: *An immunoinformatics approach* BioRxiv.
4. Pieczenik G (1994) The theory of genotypic selection: predicting the direction of evolution as a consequence of G: U base pairing and the existence of non-pathogenic strains of HIV-1. *Biochem Mol Biol Int* 32: 879-887.
5. Dawkins R (2005) *The Ancestor's Tale*, Boston: Mariner Books. p. 195.
6. Pieczenik G (1999) Method and means for sorting and identifying biological information. United States Patent and Trademark Office 5,866,363.
7. New England Biolabs (2020) *Manual- Phage Display Peptide Library*.
8. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217: 624-626.
9. King JL, Jukes TH (1969) Non-darwinian evolution. *Science* 164: 788-798.