Prediction of the effectiveness of COVID-19 vaccine candidates [version 1; peer review: 1 approved, 1 approved with reservations]

Veljko Veljkovic\textsuperscript{1}, Vladimir Perovic\textsuperscript{2}, Slobodan Paessler\textsuperscript{3}

\textsuperscript{1}Biomed Protection, Galveston, TX, 77550, USA
\textsuperscript{2}Laboratory for Bioinformatics and Computational Chemistry, Institute of Nuclear Sciences VINCA, University of Belgrade, Belgrade, 11000, Serbia
\textsuperscript{3}Department of Pathology, Galveston National Laboratory, University of Texas Medical Branch, Galveston, TX, 77555, USA

Abstract
A safe and effective vaccine is urgently needed to bring the current SARS-CoV-2 pandemic under control. The spike protein (SP) of SARS-CoV-2 represents the principal target for most vaccines currently under development. This protein is highly conserved indicating that vaccine based on this antigen will be efficient against all currently circulating SARS-CoV-2 strains. The present analysis of SP suggests that mutation D614G could significantly decrease the effectiveness of the COVID-19 vaccine through modulation of the interaction between SARS-CoV-2 and its principal receptor ACE2.

Keywords
COVID-19, SARS-CoV-2, mutations, vaccine effectiveness

This article is included in the Disease Outbreaks gateway.

This article is included in the Coronavirus collection.
Introduction
SARS-CoV-2 is the novel highly infectious human coronavirus which by May 2020 has infected 3 million and killed more than 200,000 people. Until immunity is induced in large populations throughout the world this coronavirus is likely to become endemic seasonally causing the coronavirus disease 2019 (COVID-19) in millions worldwide. Scientists and drug companies around the world are working hard to develop a vaccine against the disease with at least five candidate vaccines in clinical evaluation and another 71 in preclinical evaluation. The spike protein (SP) of SARS-CoV-2 is the principal target of most these vaccine candidates. Analysis of 5,700 isolates collected between December 2019 and April 2020 revealed only one mutation which was found in more than 1% of currently circulating viruses. This finding suggests that a single vaccine based on the consensus sequence of highly conserved SP antigen should be efficacious against current global strains.

Human coronaviruses SARS-CoV and SARS-CoV-2 recognize the angiotensin converting enzyme 2 (ACE2) as the natural receptor but present a distinct binding interface to ACE2 and a different network of residue-residue contacts. SARS-CoV and SARS-CoV-2 have comparable binding affinity but the SARS-CoV–ACE2 complex contains a higher number of contacts, a larger interface area, and decreased interface residue fluctuations relative to SARS-CoV. These suggest that the receptor binding site (RBD)–ACE2 interface of SARS-CoV-2 resembles some properties of antibody-antigen interactions, which allow the accelerated evolution of spike protein (SP) binding to the ACE2 receptor, similar to the rapid evolution along the antibody-antigen affinity maturation process. This opens the question about the effectiveness of the vaccine against SARS-CoV with mutations, which modulate its interaction with the receptor.

Previously, a novel bioinformatics approach for assessment of the effectiveness of the seasonal influenza vaccine was proposed. This approach, which is based on electronic biology, was successfully used for the prediction of the influenza vaccine effectiveness for two successive flu seasons. Here, this approach was used for the assessment of the effectiveness of the COVID-19 vaccine. This analysis showed that mutation D614G, which spreads globally, and which is present in >50% of all circulating viruses, could significantly decrease the effectiveness of the COVID-19 vaccine.

Methods
Viruses
We analyzed the subunit 1 of S protein (SP1) from SARS-CoV-2 with following mutations:

V367F, G219K, M153T, Q409E, R408I, A435S, N354D, D364Y, H655I, V615I, Q239K, Y28N, T29I, H49Y, L54F, N74K, D111N, F157L, G181V, S221W, S247R, A348T, G476S, V483A, H49Y, L54F, N74K, V367F, G219K, M153T, Q409E, R408I, A435S, N354D, D364Y, H655I, V615I, Q239K, Y28N, T29I, H49Y, L54F, N74K, D111N, F157L, G181V, S221W, S247R, A348T, G476S, V483A, H519Q, A520S, D614G.

The SARS-CoV-2 S protein reference sequence (GenPept accession YP_009724390) is used in the analysis as the wild type (WT).

The informational spectrum method
The informational spectrum method (ISM), a virtual spectroscopy method, is developed for a fast and simple structure analysis of proteins and their functionally important domains. Physical and mathematical basis of ISM is described in detail elsewhere and here the method will be presented only briefly.

A sequence of N amino acid residues is represented as a linear array of N terms, with each term given a weight. The weight assigned to a residue is the electron-ion interaction potential (EIIP), determining the electronic properties of amino acids, which are responsible for their intermolecular interactions. In this way, the alphabetic code is transformed into a sequence of numbers. The signal obtained is then decomposed in periodical function by Fourier transformation. Thus, the initial information defined by the sequence of amino acids can now be presented in the form of an informational spectrum (IS), representing the series of frequencies and corresponding amplitudes. The IS frequencies correspond to the distribution of structural motifs with defined physicochemical characteristics determining long-range interaction properties of the protein.

Phylogenetic analysis
Phylogenetic analysis of SARS-CoV-2 can be performed with the ISM-based phylogenetic algorithm ISTREE, which was previously described in detail. This phylogenetic approach that allows the assessment of the biological effect of mutations was previously applied in the analysis of influenza viruses and Zika virus. Figure 1 gives the schematic presentation of this algorithm. Here, we used an ISM distance measure defined on the specific frequency F = 0.257 which characterizes the interaction between SP1 and the ACE2 receptor.

For the development of the conventional phylogenetic tree, based on multiple sequence alignment (MSA), we used the MEGA software package. For the MSA calculation of sequences, the MUSCLE algorithm of MEGA 5 software was applied.
Results and discussion

In Figure 2a, we show the ISM-based phylogenetic tree for mutant SP1 from SARS-CoV-2. These SP1 are grouped in two separate clusters, A and B, that, according to the informational spectrum (IS) concept, have different interacting and immunological profiles. There is no difference between the analyzed SP1 in the homology-based phylogenetic tree (Figure 2b). Presented results suggest that most of the analyzed mutants (cluster A) will interact with ACE2 in a similar way as the WT virus. Only five mutations V367F, R408I, G476S, D111N, and D614G (cluster B) could significantly affect the interaction of SP1 with ACE2. Three of these mutations (V367F, R408I, G476S) are located in the RBD of SP.

The previous ISM-based phylogenetic analysis of hemagglutinins from seasonal influenza viruses showed that viruses which are in the same cluster with the vaccine virus are responsive to the seasonal flu vaccine and those viruses which are grouped in the separate clusters are resistant to the vaccine. This finding served as a base for the accurate prediction of the efficacy of the flu vaccine several months before the start of the flu season. The same approach was also used for the design of antigens for the vaccine against ZIKA virus. By analogy, it could be expected that SARS-CoV-2 with mutations V367F, R407I, G476S, D111N and D614G in SP are resistant to the COVID-19 vaccines that are based on the WT SARS-CoV-2.

The D614G mutation frequency is much higher than the others. This mutation was found in 55% of sequences sampled globally as of April 10, 2020. The next most frequent mutation is found in only 0.87% of these sequences. A virus carrying the D614G mutation, which had not then been observed among sequences from China, was transmitted in January 2020 to Germany and became dominant in Europe and then globally within two months (found in 55% of sequences).

The D614G mutation is located in the middle of one of three epitopes of SP1. This change of large acidic residue D into small hydrophobic residue G in the middle of the epitope would compromise the binding affinity to antibodies elicited by vaccines with WT SP.

Conclusions

The present results suggest that the highly prevalent D614G mutation in SP, although located outside of the receptor binding site, could decrease the efficacy of the vaccine by modulation
of the interaction of SARS-CoV with the ACE2 receptor. In addition, this mutation may cause antigenic drift, resulting in vaccine mismatches which additionally could affect the efficacy of the vaccine. These possible obstacles should be considered for the future development of COVID-19 vaccines. The further ISM-based monitoring of the evolution of SARS-CoV-2 is important for identification of other mutations, which could affect the effectiveness of vaccines against this virus.

Data availability
All data underlying the results are available as part of the article and no additional source data are required.

References

1. Dearlove B, Lewitus E, Bai H, et al.: A SARS-CoV-2 vaccine candidate would likely match all currently circulating strains. bioRxiv preprint. 2020. Publisher Full Text
2. Shang J, Ye G, Shi K, et al.: Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020. PubMed Abstract | Publisher Full Text
3. Breman JG: The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. bioRxiv preprint. 2020. Publisher Full Text
4. Veljkovic V, Paessler S, Glisic S, et al.: Evolution of 2014/15 H3N2 influenza viruses circulating in US: consequences for vaccine effectiveness and possible new pandemic. Front Microbiol. 2015; 6: 1456. PubMed Abstract | Publisher Full Text | Free Full Text
5. Paessler S, Veljkovic V: Prediction of influenza vaccine effectiveness for the influenza season 2017/18 in the US [version 1; peer review: 2 approved]. F1000Res. 2017; 6: 2067. PubMed Abstract | Publisher Full Text | Free Full Text
6. Paessler S, Veljkovic V: Using electronic biology based platform to predict flu vaccine efficacy for 2018/2019 [version 2; peer review: 2 approved, 1 approved with reservations]. F1000Res. 2018; 7: 298. PubMed Abstract | Publisher Full Text | Free Full Text
7. Mutation sites in the Spike (S) protein: China National Center for Bioinformation, 2019 Novel Coronavirus Resource (2019nCoVR). Reference Source
8. Lokman SM, Rashiduzzaman Md, Salauddin A, et al.: Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. bioRxiv preprint. 2020. Publisher Full Text
9. Veljkovic V, Cusco I, Dimitrijevic B, et al.: Is it possible to analyze DNA and protein sequences by the methods of digital signal processing? IEEE Trans Biomed Eng. 1985; 32(5): 337–41. PubMed Abstract | Publisher Full Text
10. Veljkovic V, Slavic I: Simple general-model pseudopotential. Phys Rev Let. 1972; 29: 105. Publisher Full Text
11. Veljkovic V: The dependence of the Fermi energy on the atomic number. Physics Letters A. 1973; 46(1): 41–42. Publisher Full Text
12. Veljkovic V: A theoretical approach to preselection of carcinogens and chemical carcinogenesis. Gordon & Breach, New York, 1980. Reference Source
13. Pernovic VR, Muller CP, Niman HL, et al.: Novel phylogenetic algorithm to monitor human tropism in Egyptian H5N1-HPAIV reveals evolution toward efficient human-to-human transmission. PLoS One. 2013; 8(4): e61572. PubMed Abstract | Publisher Full Text | Free Full Text
14. Veljkovic V, Veljkovic N, Muller CP, et al.: Characterization of conserved properties of hemagglutinin of H5N1 and human influenza viruses: possible consequences for therapy and infection control. BMC Struct Biol. 2009; 9: 21. PubMed Abstract | Publisher Full Text | Free Full Text
15. Veljkovic V, Niman HL, Glisic S, et al.: Identification of hemagglutinin structural domain and polymorphisms which may modulate swine H1N1 interactions with human receptor, BMC Struct Biol. 2009; 9: 62. PubMed Abstract | Publisher Full Text | Free Full Text
16. Veljkovic V, Veljkovic N, Paessler S, et al.: Predicted enhanced human propensity of current avian-Like H1N1 swine influenza virus from China. PLoS One. 2016; 11(1): e0156515. PubMed Abstract | Publisher Full Text | Free Full Text
17. Veljkovic V, Glisic S, Muller CP, et al.: In silico analysis suggests interaction between Ebola virus and the extracellular matrix. Front Microbiol. 2015; 6: 135. PubMed Abstract | Publisher Full Text | Free Full Text
18. Paessler S, Veljkovic V: Vaccine with reduced enhancement of viral infection. United States Patent 10322171. 2019. Reference Source
19. Jones RF, Balint J, Rice A, et al.: Compositions and methods for flavivirus vaccination. Patent WO 2018/014006 Al. 2018. Reference Source
20. Veljkovic V, Vergara-Alert J, Segales J, et al.: Use of the informational spectrum methodology for rapid biological analysis of the novel coronavirus 2019-nCoV: prediction of potential receptor, natural reservoir, tropism and therapeutic/vaccine target [version 3; peer review: 2 approved]. F1000Res. 2020; 9: 52. Publisher Full Text
21. Tamura K, Peterson D, Peterson N, et al.: MEGA6: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28(10): 2731–2738. PubMed Abstract | Publisher Full Text | Free Full Text
22. Koyama T, Weeraratne D, Snowdon JL, et al.: Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. Preprints (www.preprints.org). Publisher Full Text
The authors are trying to find out important residues of spike protein interacting with ACE2 using the ISM method.

- The application of the methods have not been validated with wide range of proteins comparing with experimental data. Therefore, it is not easy to evaluate the effectiveness of the method. Experimental data is available in the following paper:
  Chan, K.K., et al., Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science, 2020: p. eabc0870.¹

I would like to see how ISM is agreement with the data.

- Secondly, the ISM method is attempting to identify the residue impacting interactions with ACE2. However, for vaccines, neutralizing antibodies attached to spike protein to inhibit interaction or other process. Therefore, direct interaction with ACE2 does not necessarily mean key residue for vaccine. Rather, antibodies from B-cell must bind to the residues and the residues identified by the authors are not necessarily in the B-cell epitopes. Therefore, those residues are irrelevant to vaccines.

References
1. Chan KK, Dorosky D, Sharma P, Abbasi SA, et al.: Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science. 2020. PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** genomics, oncology, immunology, stem cell, virology, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 14 Aug 2020

**Veljko Veljkovic**, Biomed Protection, Galveston, USA

The Reviewer’s remark:

“The application of the methods have not been validated with wide range of proteins comparing with experimental data. Therefore, it is not easy to evaluate the effectiveness of the method.”

Response

Reviewer can find several articles in PubMed where ISM prediction is experimentally demonstrated (e.g. doi: 10.1038/s41598-018-20185-8. doi: 10.1038/srep11434). Based on previous work we believe the ISM does not need any additional proof and can be used to estimate efficacy as it is. Moreover, the Reviewer overlooked that the ISM analysis concerns the long-range interactions between molecules (interactions at distances between 5 and 1000 A) and not their chemical binding (the interactions at distances <5A). The ACE2 example suggested by the Reviewer is not suitable for the ISM analysis because it concerns the improvement of the chemical binding of SP to this receptor.

The Reviewer’s remark:

“Secondly, the ISM method is attempting to identify the residue impacting interactions with ACE2. However, for vaccines, neutralizing antibodies attached to spike protein to inhibit interaction or other process. Therefore, direct interaction with ACE2 does not necessarily mean key residue for vaccine. Rather, antibodies from B-cell must bind to the residues and the residues identified by the authors are not necessarily in the B-cell epitopes. Therefore, those residues are irrelevant to vaccines”
Response

Presented ISM results indicate that the mutation D614G modulate the recognition and targeting between antibodies and SP antigen and not their “direct interaction” (chemical binding). If the mutation prevents the efficient recognition between antigen and an antibody, the neutralization will not be efficient even if the antibody would efficiently bind to the antigen. In other words, the antibody will not efficiently “find” its target in vivo.

**Competing Interests:** No competing interests were disclosed.

---

**Reviewer Report 15 June 2020**

[https://doi.org/10.5256/f1000research.26334.r64327](https://doi.org/10.5256/f1000research.26334.r64327)

© 2020 Pongor S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

![✅ Sándor Pongor](Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary)

- This work highlights an interesting aspect of the COVID spike protein that may be useful for vaccine developers.

  - Sequences of spike protein mutants are compared with the information spectrum method developed by the authors and it is found that the clustering of spike protein mutants allows one to draw conclusions on the expected effectiveness of the vaccines.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes
Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com