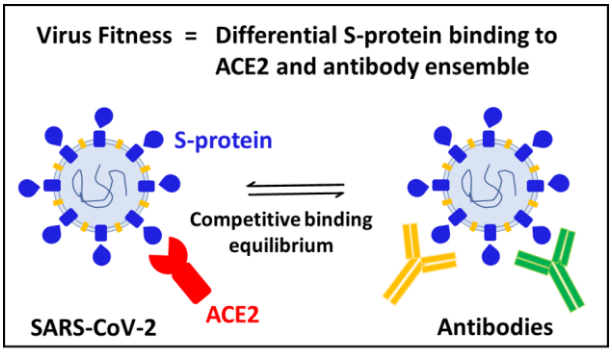
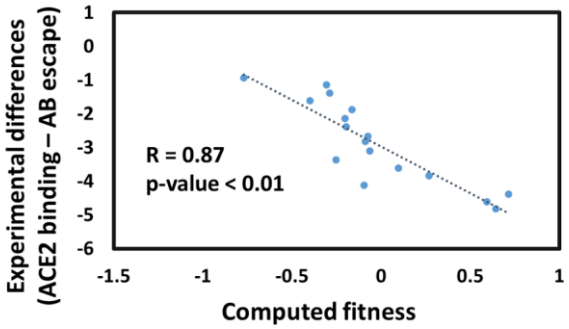
Computing Viral Fitness: Towards a structure-based approach

Rukmankesh Mehra

Department of Chemistry, Indian Institute of Technology Bhilai, Bhilai, Durg 491001, Chhattisgarh, India. E-mail: rukmankesh@iitbhilai.ac.in.





Computing the impact of new emerging virus mutations is of major interest for understanding the evolutionary forces of the pathogen and for surveillance purposes. The SARS-CoV-2 spike-protein (S-protein) is central to the current vaccine development efforts. This protein binds to human ACE2 receptors as a critical step in host cell infection, but also to human antibodies that neutralizes the virus and prevents interaction with ACE2. We proposed that this host-virion interaction can be viewed as a competitive binding situation where the virion seeks to enter the human cell via ACE2 binding before the S-protein is bound to circulating antibodies [1-5]. If so, the thermochemical selectivity (differential binding affinity to ACE2 vs. the ensemble of circulating antibodies) defines the fraction of virions that infects cells before they are neutralized by antibodies. Combining the two binding events seems important as many individual mutations generically either strengthen or weaken binding to many proteins: Stronger binding to ACE2 is not a good fitness metric if the mutant at the same time leads to equally stronger binding to antibodies. Here, we defined a simple *fitness model* of SARS-CoV-2 based on the two binding properties of the S-protein to its human host cell-surface receptor ACE2 and a representative ensemble of diverse antibodies circulating in the human population [1,2]. We implemented the model using structure-based computation of all possible mutation effects averaged over 10 ACE2 complexes and 10 antibody complexes of the S-protein (~3,80,000 computed mutations) and study all possible mutations in the S-protein to provide a full heat map of estimated fitness effects. The use of many protein structures ensures much more robust estimates of the effects, and the selectivity model also takes advantage of systematic error cancellation by considering "differences between differences" in estimated binding affinities (changes in ACE2–AB selectivity), which is another novel advantage compared to computational or experimental fitness estimates based on just one of the binding events. The approach was validated by correlating our computed data with diverse experimental binding/escape data of ACE2 and antibodies [1-2]. Our computed fitness correlated well with the experimental fitness [1]. The method and our results should be of substantial interest, with a potential for use in surveillance as an early estimator of the potential concern of new arising mutations.

[1] S. Thakur, K. P. Kepp, R. Mehra, *J. Struct. Biol.*, **2023**, *215(4)*, 108042.
[2] S. Thakur, R.K. Verma, K. P. Kepp, R. Mehra, *J. Mol. Graph. Model.*, **2023**, 119, 108379.
[3] R. Mehra, K. P. Kepp, *ACS Infect. Dis.*, **2022**, *8(1)*, 29-58.
[4] R. Mehra, K. P. Kepp, *Eur. Biophys. J.*, **2022**, *51(7-8)*, 555-568.
[5] K.T. Bæk, R. Mehra , K. P. Kepp, *Mol. Cell. Biochem.*, **2023**, *478(6)*,1269-1280.