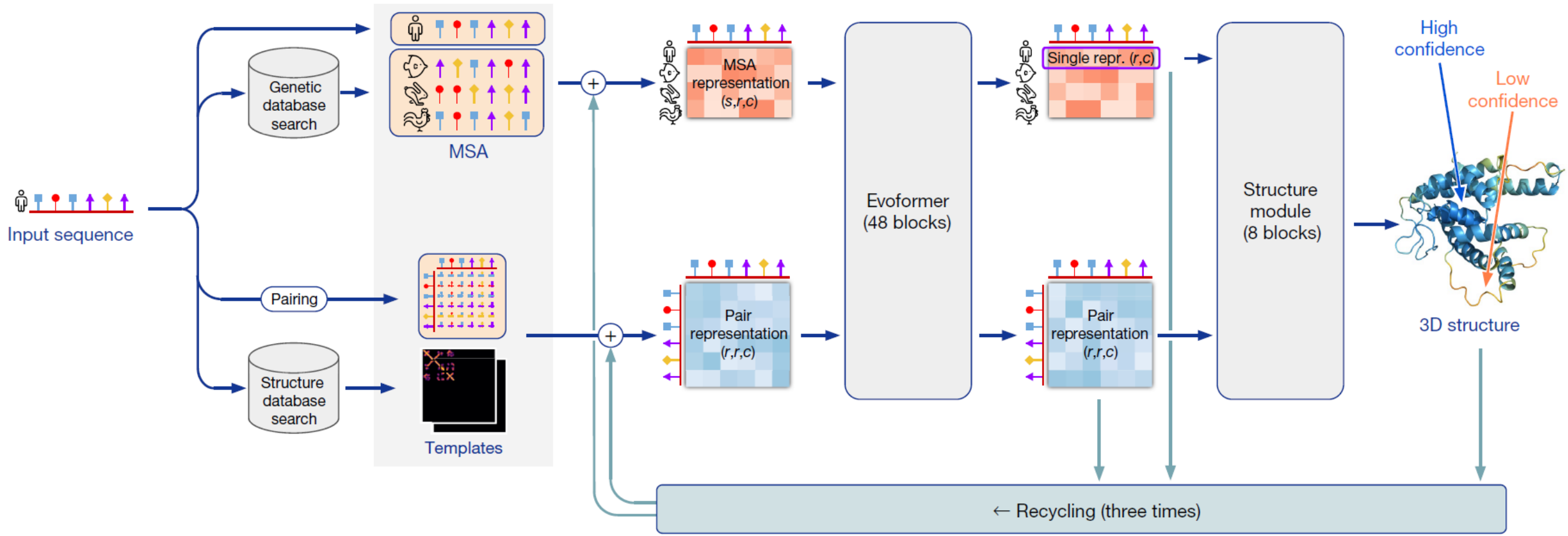
Exploring SARS-CoV-2 receptor binding domain variants

Ákos Gellért, Oz Kilim, Anikó Mentes, Balázs Pál, István Csabai

Project Team members introduction

Oz Kilim – Machine learning specialist
Anikó Mentes – Bioinformatics, big data handling
Balázs Pál – IT problem solving
István Csabai – Team leader, scientific think tank
Ákos Gellért – AlphaFold running, data analysis

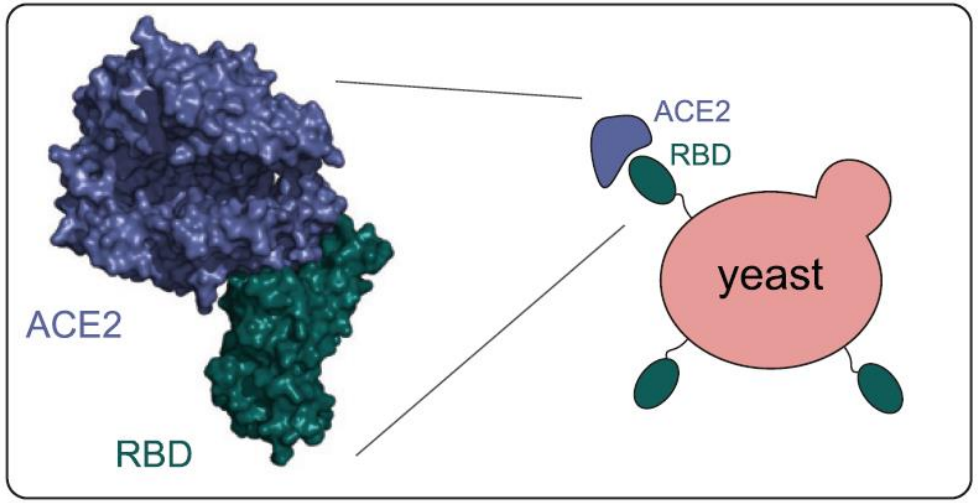
AlphaFold model architecture



J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nat. 2021 5967873*, vol. 596, no. 7873, pp. 583–589, Jul. 2021, doi: 10.1038/s41586-021-03819-2.

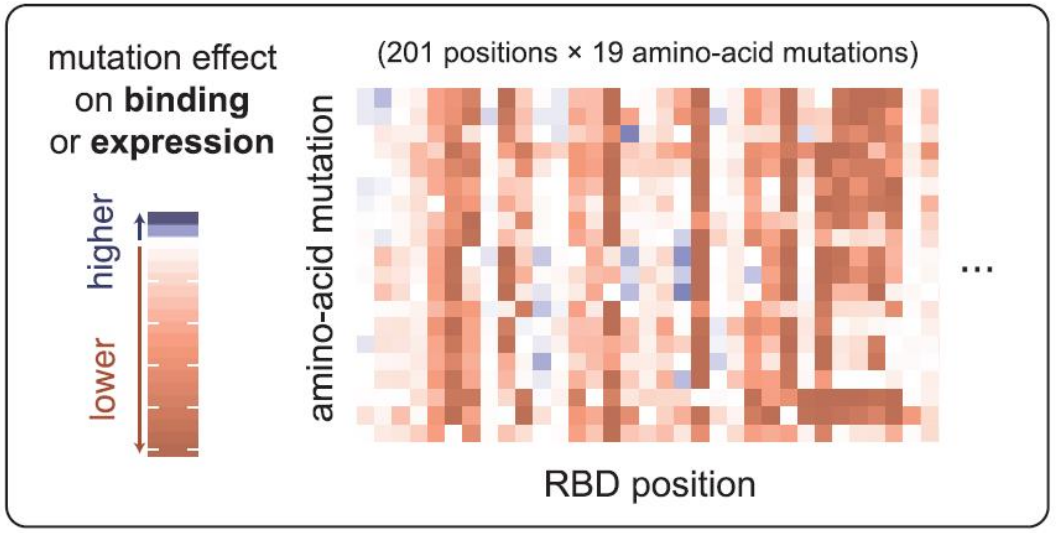
Deep Mutational Scanning

SARS-CoV-2 receptor binding domain (RBD) binds human ACE2 receptor

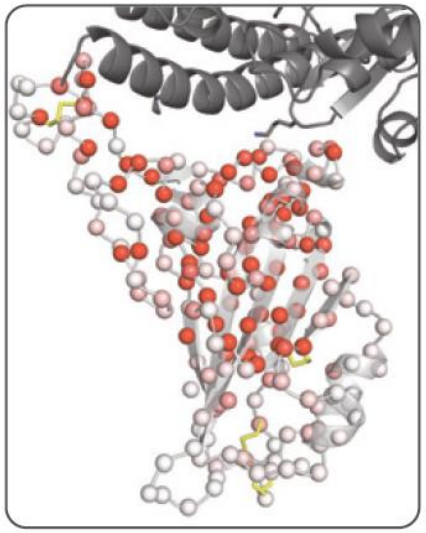


constructed library of all amino acid mutations in SARS-CoV-2 RBD

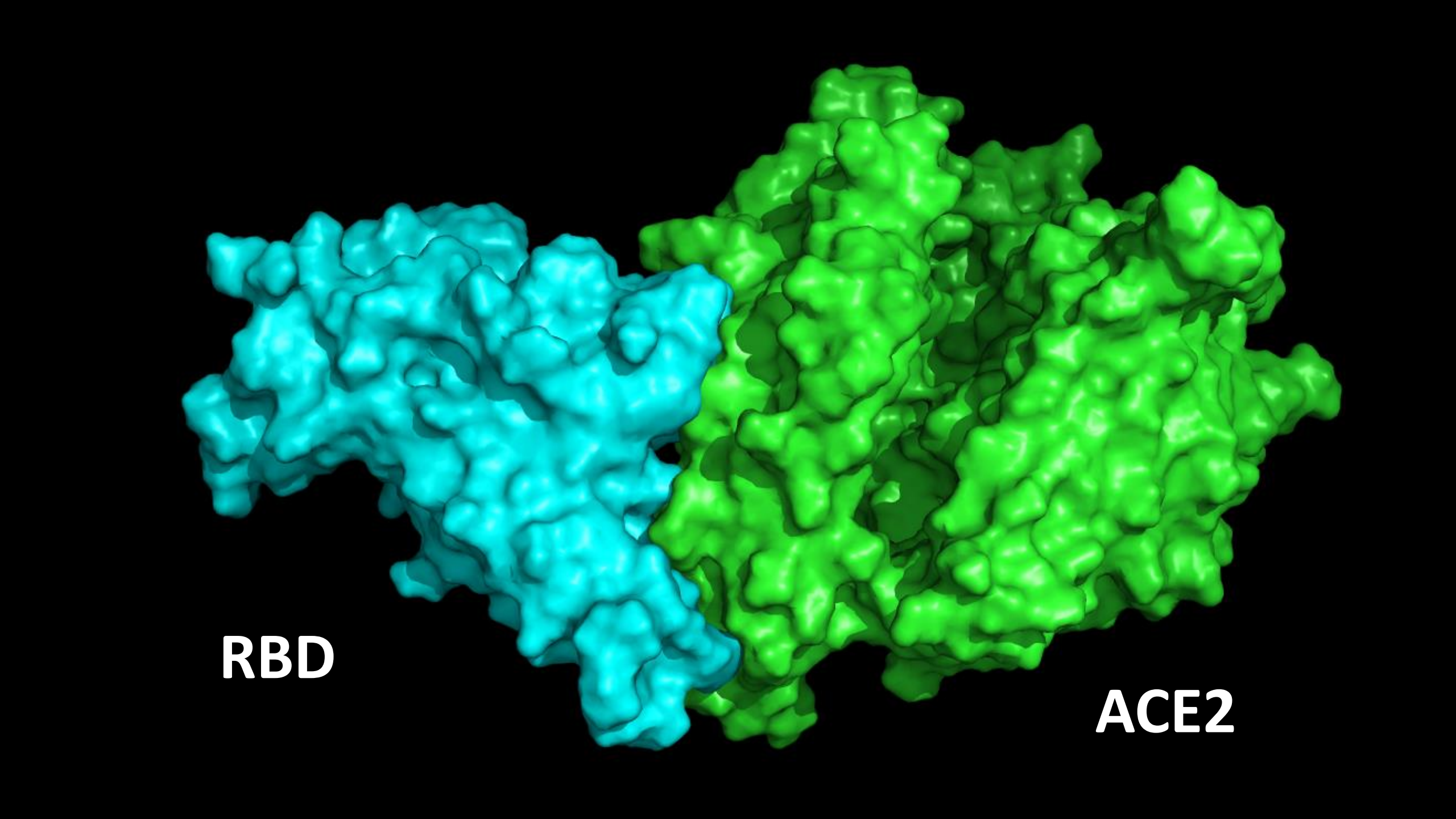
determined effects of all mutations on ACE2-binding affinity and RBD folding

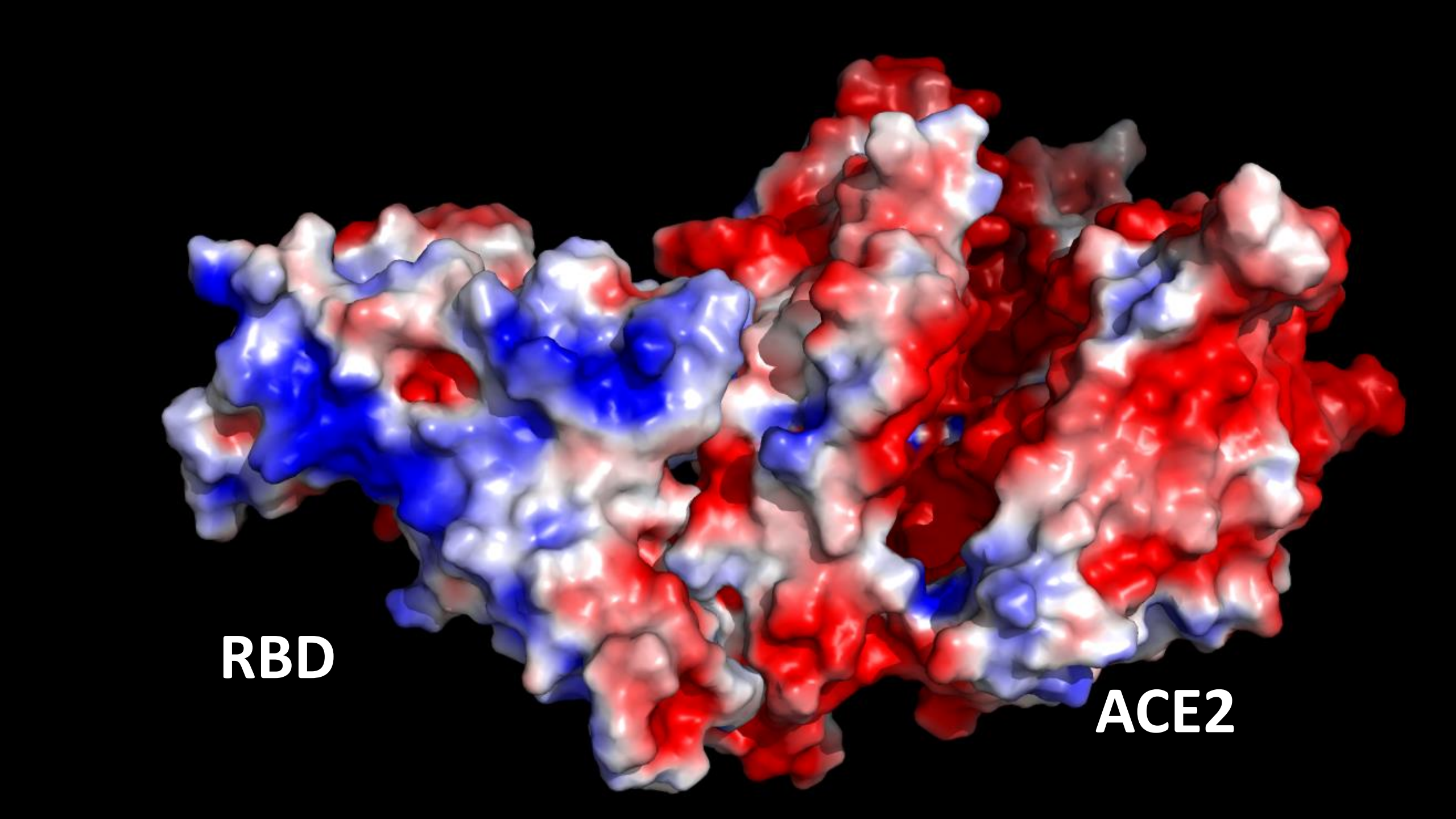


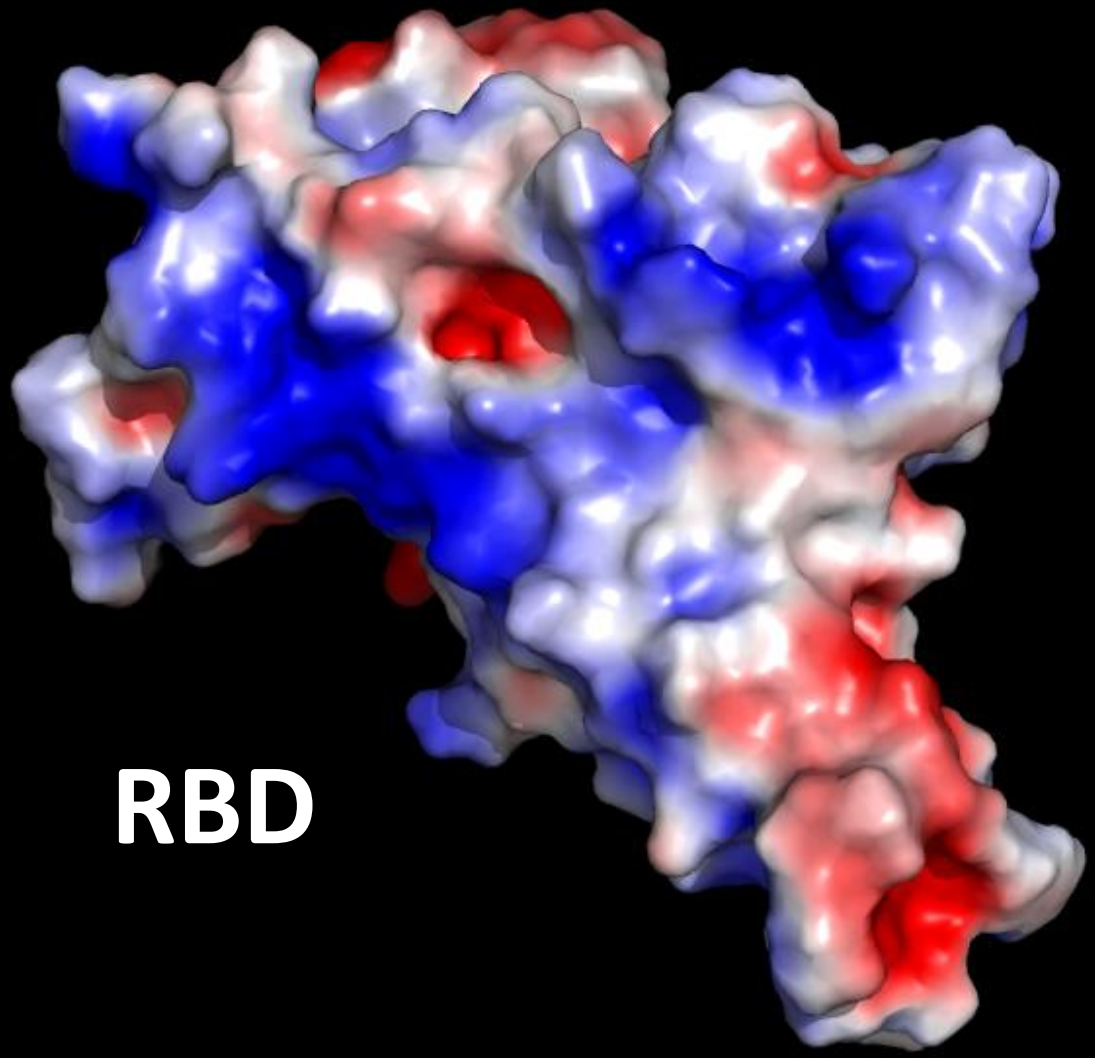
RBD structure-function relationship



T. N. Starr *et al.*, "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding," *Cell*, vol. 182, no. 5, pp. 1295-1310.e20, Sep. 2020, doi: 10.1016/J.CELL.2020.08.012.

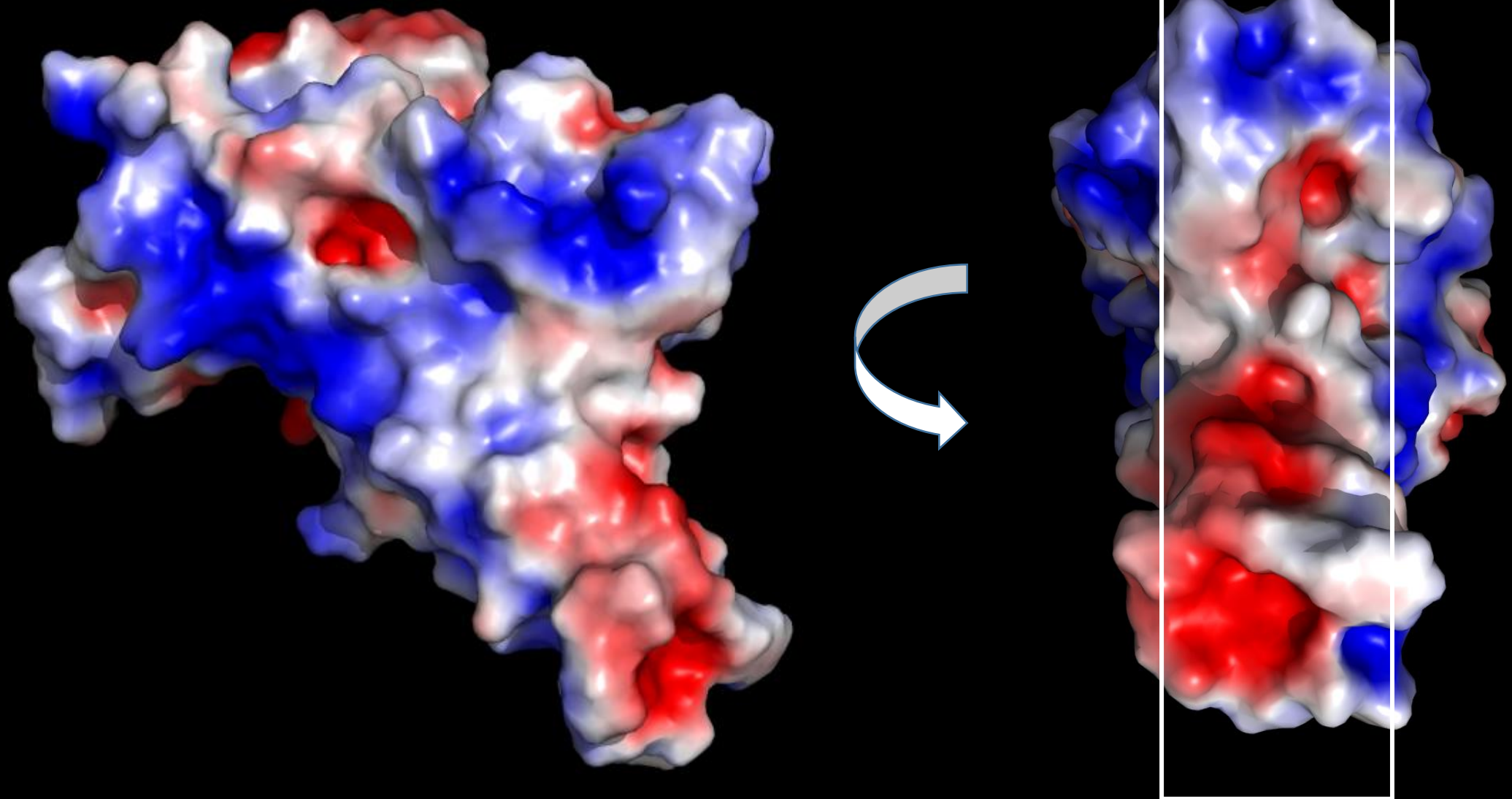

RBD
ACE2

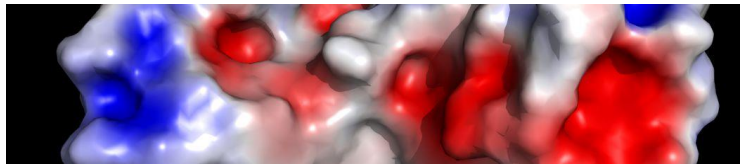
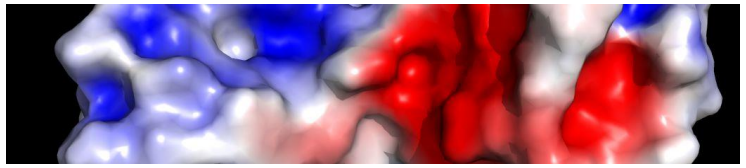
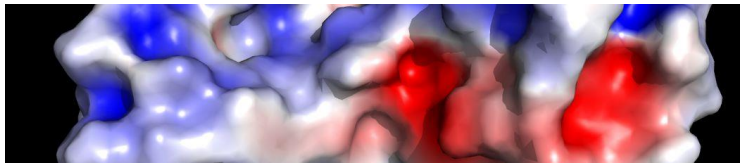
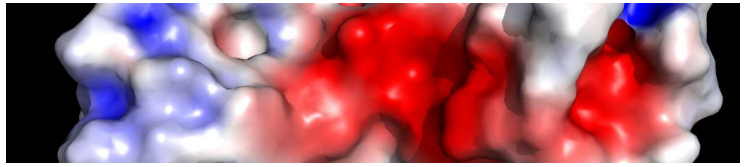
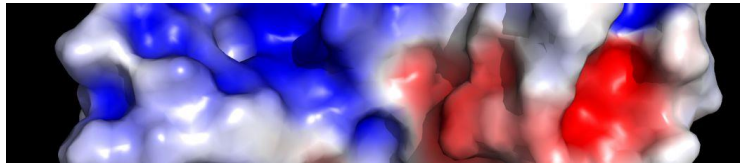
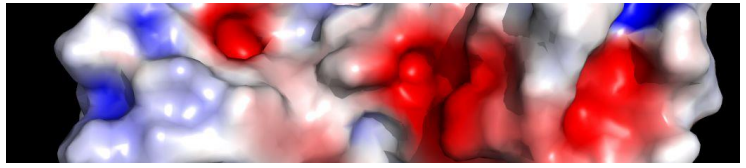

RBD
ACE2



RBD

Binding interface



		Change in ACE2 binding
RBD_Wuh_IF		
RBD_Y489E_IF		-3.55
RBD_Y489R_IF		-2.83
RBD_Q493E_IF		-0.98
RBD_Q493R_IF		-0.30
RBD_Y505E_IF		-3.73
RBD_Y505R_IF		-3.54

Expected results

Structure database

- Single mutant structural database generated with AlphaFold2 for RBD variants: Wuhan, alpha, beta, delta, eta and omicrons

already done

Develop machine learning methods

- Using electrostatic and shape descriptors from RBD structures correlated with experimental ACE2 binding affinity

- The previous method combined with protein-protein docking calculations

The main goal is to try to develop a machine learning based method that can produce deep mutational scanning experimental data *in-silico.*

So there should be no need to perform expensive molecular biology experiments. ☺ But for now, it is still a dream.

Technical details

Ampere01 @ Wigner Research Centre for Physics
256 CPUs - AMD EPYC 7742 64-Core Processor
8 nVIDIA A100 80GB GPU cards

For AlphaFold database access time acceleration an 5 TB SSD is used.

Running time for one RBD model is about 15 to 20 minutes.
10 parallel run per GPU card.

Thanks for my colleagues and for the GPU Team at Wigner