

Comparing Receptor Binding Properties of SARS-CoV-2 and of SARS-CoV Virus by Using Unsupervised Machine Learning Models

Cao Phuong Cong, Hien T. T. Lai, Ly H. Nguyen, and Toan T. Nguyen[†]

Key Laboratory for Multiscale Simulation of Complex Systems, and Faculty of Physics, University of Science, Vietnam National University, 334 Nguyen Trai Street, Thanh Xuan, Hanoi 11400, Vietnam

Email: toannt@hus.edu.vn

ID: INV10

Invited Talk

Aug. 24, 11.00 - 11.30

This work continues our recent molecular dynamics investigation of the three systems of the human ACE2 receptor interacting with the viral RBDs of SARS-CoV virus and two variants of SARS-CoV-2 viruses. The simulations are extended and analyzed using unsupervised machine learning models to give complementary descriptions of hidden features of the viral binding mechanism. Specifically, the principal component analysis (PCA) and the variational autoencoder (VAE) models are employed, both are classified as dimensionality reduction approaches with different focuses. The results support the molecular dynamics results that the two variants of SARS-CoV-2 bind stronger and more stable to the human ACE2 receptor than SARS-CoV virus does. Moreover, stronger bindings affect the structure of the human receptor, making it fluctuate more, a sensitive feature which is hard to detect using standard analyses. Unexpectedly, it is found that the VAE model can learn and arrange randomly shuffled protein structures obtained from molecular dynamics in time order in the latent space representation. This result potentially has promising application in computational biomolecules. One could use this VAE model to jump forward in time during a molecular dynamic simulation, and to enhance the sampling of protein configuration space.

Keywords: Coronaviruses, human ACE2, unsupervised machine learnings, enhanced sampling, molecular dynamics, variable autoencoder