Development of Conformational Descriptors for Organic Materials Informatics

(*National Institute of Advanced Industrial Science and Technology (AIST)*) ○Hiroshi Izumi

Keywords: Conformation; Deep Neural Network; Maximal Common Substructure; COVID-19; IUPAC Nomenclature

Design of organic materials using fingerprints of chemical structural formula has often restrictions. Big data of properly selected maximal common substructures (MCSs) may avoid the heavy loads of theoretical calculations and may be available for the design of new functional molecules. For this purpose, the codification techniques of conformations for comparison of MCSs have been developed. In this presentation, a conformational descriptor, which is called protein supersecondary structure code (SSSC), and the application of SSSC to a deep neural network-based software for the prediction of conformational variability are described.

SSSC was developed that uses the concept of Ramachandran plot data with ω angles, and the specification of positions of MCSs in a protein derived from a fuzzy search of structural code homology using template patterns, represented as conformational codes 3a5c4a (α-helix-type conformation) and 6c4a4a (β-sheet-type conformation), to describe supersecondary structural motifs and their conformations.[1,2] SSSC is transcribed using the letters H, S, T, and D to refer to an α-helix-type conformation, a β-sheet-type conformation, other-type conformations, and disordered residues or the C-terminus, respectively. The predicted conformational variability of the mutation sites for SARS-CoV-2 spike proteins correlates with the transmissibility with high expression and the neutralization escape ability well.[3] Especially, the predicted conformational variability of FURIN sites for α, δ, and o strains are convenient for the binding of S1 (C-terminus; XSXRRAR) to neuropilin-1 (Fig. 1). It is suggested that the conformational descriptors of organic molecules are necessary for the design of organic materials with the high energy-barrier of conformational change.

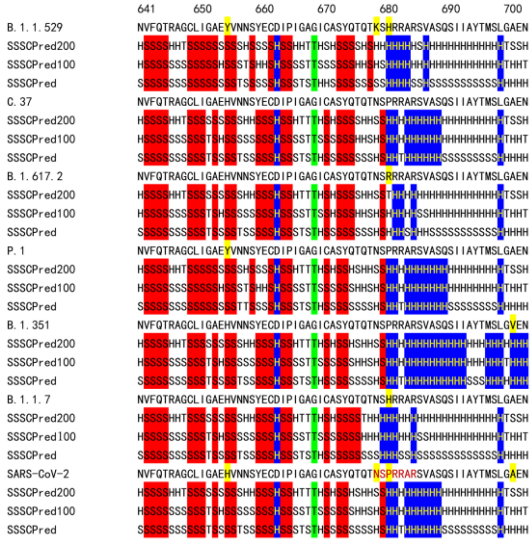

Fig 1. Conformational variability prediction of multiple mutation strains.

1) H. Izumi, *Methods Mol. Biol.* **2019**, *1958*, 329. 2) H. Izumi, A. Wakisaka, L. A. Nafie, R. K. Dukor, *J. Chem. Inf. Model.* **2013**, *53*, 584. 3) H. Izumi, L. A. Nafie, R. K. Dukor, *ACS Omega* **2021**, *6*, 19323.

This work was partly supported by JSPS KAKENHI Grant Number JP19K05431.