

Prediction and Design of Cyclodextrin Inclusion Complexes formation with Machine Learning-based Strategies

Yiming Ma¹, Yue Niu¹, Huaiyu Yang², Jiayu Dai¹, jiawei lin¹, Huiqi Wang¹, Songgu Wu¹, Qiuxiang Yin¹, Ling Zhou¹, and Junbo Gong¹

¹Tianjin University

²Loughborough University

May 6, 2022

Abstract

This work aims to develop multi-purpose machine learning (ML)-based cyclodextrin inclusion complexes (ICs) formation predicting strategies in aqueous solution to replace traditional experimental approaches. A balanced dataset of drug relevant molecules was constructed with experimental verifications. Three ML models (artificial neural network, support vector machine, and logistic regression) were established and optimized for ICs formation prediction. In order to provide more reliable approaches for different prediction requirements, ML-based linear strategy, recall-first strategy, and precision-first strategy were further established based on the ML models to pursue the maximum recall or precision values. It has also been proved that the proposed recall-first strategy finds all positive samples as much as possible to avoid missing in prediction, and the precision-first strategy finds positive samples accurately to reduce the number of validation experiments. The ML-based prediction strategies for ICs formation were first established in this work and showed high accuracy and reliability.

Hosted file

Manuscript.docx available at <https://authorea.com/users/480991/articles/568199-prediction-and-design-of-cyclodextrin-inclusion-complexes-formation-with-machine-learning-based-strategies>

