AUSTRALIAN SOCIETY FOR BIOPHYSICS

Abstract: 170P

Machine learning approaches for the discovery and optimization of oxazolidinones with therapeutic potential

McKenna E. Buckley[a,b], Santu Rana[c], Kathryn E. Fairfull-Smith[b,d], Neha S. Gandhi[a,b]

[a]Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, QLD Australia; [b]School of Chemistry and Physics, Queensland University of Technology, Brisbane, QLD, Australia; [c]Applied Artificial Intelligence Institute (A^2I^2), Deakin University, Geelong, VIC, Australia; [d]Centre for Materials Science, Queensland University of Technology, Brisbane, QLD, Australia;

Structure-based virtual screening methods, such as molecular docking, have provided a cost-effective and convenient in silico solution in the early stages of protein or RNA drug discovery. Molecular docking uses structural information to estimate receptor-ligand recognition, providing valuable information on large chemical libraries at a rapid pace.[1] However, compared to protein-ligand binding, ligand binding sites on RNA can be less deep, more polar, solvated, and conformationally flexible, which adds further complexity when predicting RNA-ligand interactions.[2] Therefore, ligand-based virtual screening methods, such as machine learning, have become an increasingly popular alternative solution. Machine learning has evolved as a critical technology in small-molecule drug discovery, with techniques based on relative molecular similarity analysis of compounds with known and unknown activity. This research aims to analyse and compare different machine learning algorithms and to identify the best predicting algorithm using a dataset of oxazolidinone class molecules with known minimum inhibitory concentration (MIC) activity. Oxazolidinones are a broad-spectrum class of synthetic antibiotics binding to the 50S ribosomal subunit of gram-positive and gram-negative bacteria. Machine learning models to predict activity were constructed using Morgan molecular fingerprints, with a dataset consisting of 530 oxazolidinones molecules with known MIC values obtained from the literature. An evaluation of several candidate algorithms on the primary dataset revealed that the support vector regressor (SVR) gave the best model, with a coefficient of determination (R2) of 0.703. These algorithms were then repeated for the various clustered sub-groups of the dataset to determine potential influences or variations in results. SVR is then used as a prediction metric for structures of unknown MIC and measured against existing literature values to compare the accuracy of its predicting power. By modifying existing algorithms and fitting them against a novel dataset, this analysis method could be used in further novel drug discovery efforts that utilise structure-based techniques.

References

1. Varela-Rial, A.; Majewski, M.; De Fabritiis, G., (2021), Structure based virtual screening: Fast and slow, *WIREs Comput Mol Sci*., **12**, 1544-1561
2. Luo, J.; Wei, W.; Waldispühl, J.; Moitessier, N., (2019), Challenges and current status of computational methods for docking small molecules to nucleic acids, *Eur J Med* Chem., **168**, 414-425