Clara MARTINEZ
Biotech 4
Sup'Biotech

Qubit Pharmaceuticals

Prediction and Understanding of hERG Inhibition using Machine Learning and Structure-Based Drug Design

The hERG channel is a voltage-gated potassium channel involved in the electrical activity of the heart. The inhibition of the hERG channel leads to a high risk of cardio toxicology and thus has become a critical issue in the pharmaceutical industry. In recent years, computational molecular models have accelerated the drug discovery process, but they have not been able to solve the binding modes of hERG blockers. This study aims to understand and predict how small molecules interfere with the channel, and its promiscuity to bind with a wide range of drugs. To predict if an unknown molecule can be a potential drug candidate and thus is not an hERG inhibitor, we have built several QSAR models applying the Decision Tree, Random Forest, and Light-GBM algorithms. The obtained models were externally validated using a "One-Cluster-Out Cross-Validation" strategy specifically developed to assess the predictive ability of the model in a closer real-world scenario. Furthermore, using structure-based drug design a docking study was conducted to gain insight into the binding modes of the hERG inhibitors. The developed QSAR models have showed good prediction accuracy (balanced accuracy exceeding 70%) for the Random Forest and Light-GBM models. However, only the Light-GBM model maintains a low % of error (under 10%) in the classification of the external validation data. Additionally, four binding modes of hERG inhibitors were discovered. All uncovered binding modes highlight the fundamental role of tyrosines (Y652) in the stabilization and binding of the drugs in the central cavity of the protein, which has been previously characterized by several studies. Nonetheless, several other amino acid residues, such as the alanines (A653), the serines (S660 and S624), and the glycines (G657), were observed to be responsible for key interactions in the binding modes.

Machine Learning, hERG, Structure-based drug design, QSAR, Docking, ADME-Tox, Decision Tree, Random Forest, Light-GBM, Cross Validation, HBDSCAN, UMAP, Cardiotoxicity