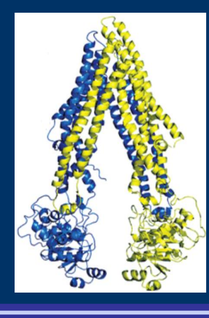
Medway School of Pharmacy

Email Ghafourian T

Evaluation of QSAR and ligand enzyme docking for the identification of ABCB1 substrates

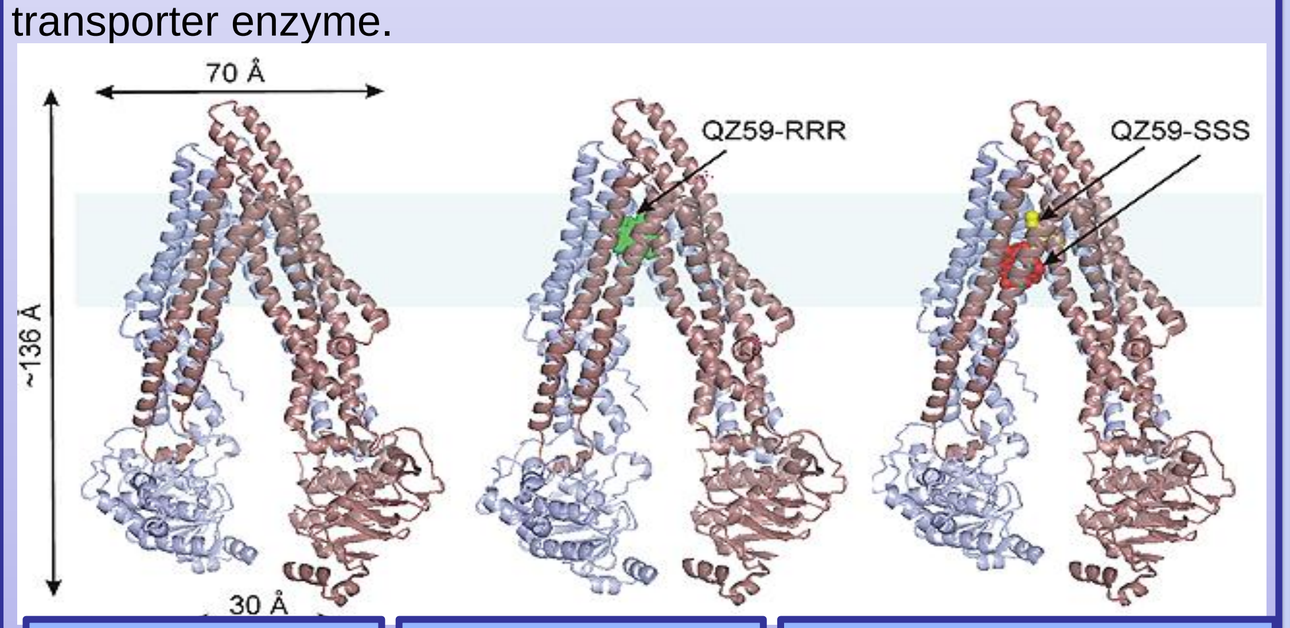
V. Osho, O. Ojo, M. Sharifi and T. Ghafourian
Medway School of Pharmacy, Chatham Maritime, Kent ME4 4TB

University of Kent



Introduction

The ABC transporter superfamily is one of the largest and abundant families of proteins. The ABC transporter P-glycoprotein (ABCB1, P-gp), a polyspecific protein has demonstrated its function as a transporter of hydrophobic drugs as well as transporting lipids, steroids and metabolic products. Its role in multidrug resistance (MDR) and pharmacokinetic profile of clinically important drug molecules has been widely recognised. Figure below shows **X-Ray crystal structures of P-glycoprotein** that are available in the Protein Data Bank (PDB) with various binding sites of this polyspecific transporter enzyme.



3G5U *without a co-crystallised ligand*	***3G60*** *co-crystallised with QZ59-RRR*	***3G61*** *co-crystallised with two molecules of QZ59-SSS*

In this study, QSAR and enzyme-ligand docking methods were explored in order to classify substrates and non-substrates of P-glycoprotein.

Methods

Dataset: The p-gp inhibition class published by Matsson et al (2009) [1] consisting of 54 inhibitors (IC_{50} <50 µM) and 69 non-inhibitors. Dataset was split randomly into a training set of 98 compounds for building models and a validation set of 25 for testing the model accuracy.

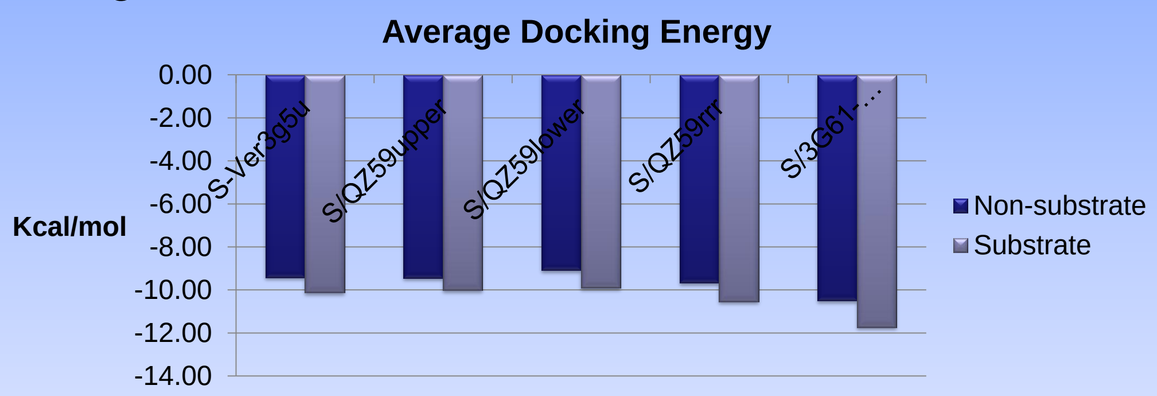
Molecular Descriptors: ACD Labs/Log D Suite Version 12.0 and Molecular Operating Environment (MOE), 2012.10 were used to calculate molecular properties.

Docking: Docking of compounds was carried out using the Dock application in MOE software. Compounds were docked into the X-ray structures of mouse P-gp 3G5U and 3G61 [2] extracted from Protein Data Bank. Docking experiments included four different binding sites on 3G5U protein [2].

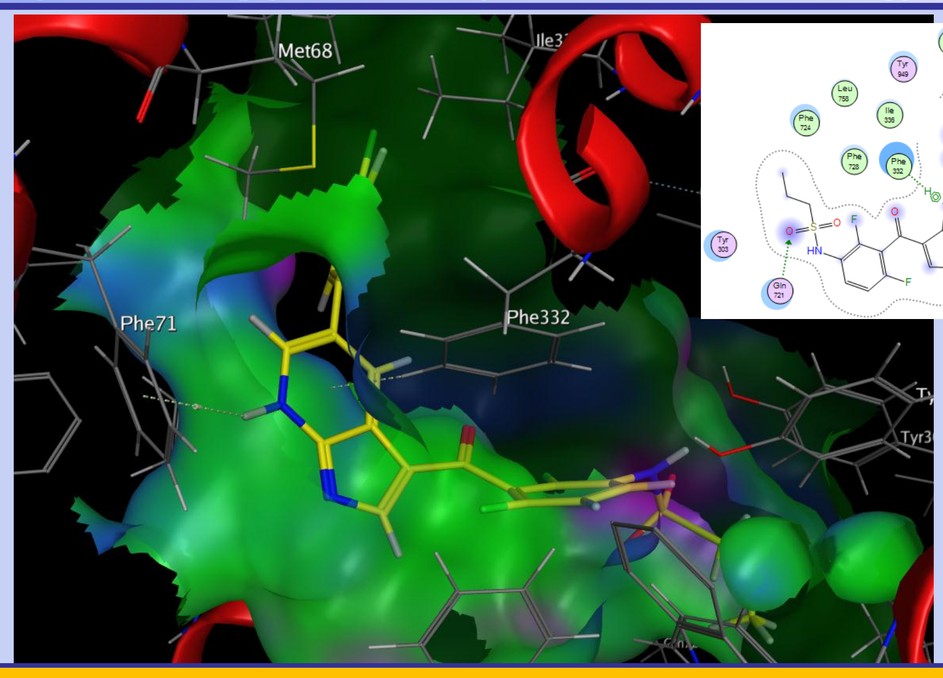
Statistical Analysis: Results from docking experiments (scores) and molecular descriptor calculations were analysed using data mining tools including CART, boosted trees and Support Vector Machine (Statistica 11.0).

Results

Docking scores



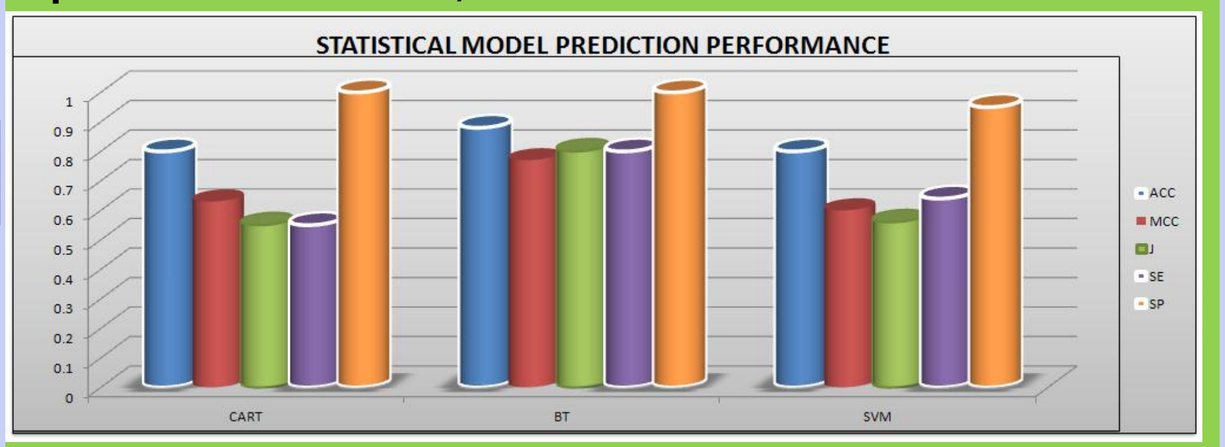
An example docking pose: Chlorprotixene docked to P-gp 3G61



Significance of docking scores using selected CART, iCART and SVM models; ACC= Accuracy; MCC= Mathews Correlation Coefficient

			Training Set		Validation Set	
Model	Parameters given	Parameters selected by analysis	Acc	MCC	Acc	MCC
CART 1	Docking scores	Docking scores at QZ59rrr and QZ59lower	0.71	0.49	0.64	0.35
CART 2	Docking scores + molecular descriptors	Various molecular descriptors	0.96	0.91	0.48	0.12
iCART 1	QZ59rrr (manually selected) + molecular descriptors	QZ59rrr, log D_2, log D_{10}	0.83	0.66	0.81	0.61
iCART 2	QZ59lower (manually selected) + molecular descriptors	QZ59lower, log D_2, log D_{10}, Q_VSA_HYD	0.85	0.71	0.81	0.61
SVM	QZ59rrr, log D_2, log D_{10}	QZ59rrr, log D_2, log D_{10}	0.76	0.5	0.81	0.65

Optimization of CART, SVM and Boosted trees



ACC= Accuracy ; MCC= Mathews Correlation Coefficient; SE= Sensitivity; SP= Specificity

Selected Models:

Model	Train set		Validation set	
	ACC	MCC	ACC	MCC
CART	0.94	0.88	0.80	0.63
BT	1.00	1.00	0.88	0.77
SVM	0.85	0.69	0.80	0.60

Conclusion

Docking performance was better using the 3G61 structure of P-gp. The most important feature for binding to P-gp was lipophilicity. Use of lipophilicity and docking scores in Support Vector Machine leads to the most accurate prediction model in comparison with the models based on docking scores only. On the other hand, optimization of CART, SVM and BT without the use of docking scores may produce similar or better results (in case of boosted trees). A model generated using BT was identified as the best model, with a prediction accuracy of 88%, Mathews correlation coefficient of 0.77 and Youden's J index of 0.80 for the test set.

References

1. Aller et al (2009). *Science*; 323(5922): p1718-22.
2. Matsson et al (2009). *Pharmaceutical Research*; 26(8): p1816-31.