P-pg Transporter Models

P-glycoprotein (P-gp) is an ATP driven efflux pump encoded by the MDR1 gene, capable of transporting a wide spectrum of chemical structures as well as different classes of drugs. Active transport by P-gp can represent a serious hurdle for pharmaceuticals as transport by P-gp has been associated with reduced bioavailability of orally administered drugs and with decreased ability of drug candidates to cross blood-tissue barriers such as the blood-brain barrier. In addition, if a drug is subject to significant P-gp efflux, its distribution, absorption and elimination could be altered by potent P-gp inhibitors leading to drug-drug interactions. Therefore, from the drug discovery and development perspective, knowledge of the transport of drug candidates by P-gp is desirable at an early stage of the drug design process.

Data

Zdrazil et al. (Mol. Inf. 31(8), pp. 599–609 (2012)) published a review of public domain data from assays of P-gp efflux activity. In this paper they developed two data sets:

- 203 compounds classified as inhibitor/non-inhibitor. This combined data from different assay types by classifying all compounds having an equal or better affinity/potency than verapamil in a certain assay as an inhibitor. Compounds with an activity value lower than verapamil were classified as non-inhibitors. We identified a number of duplicate compounds in this data set and, after removing these, 190 compounds remained.
- 198 compounds with pEC50 data from a daunorubicin transport assay in MDR CCRF vcr1000 cells. The authors note “The size of this dataset definitely satisfies QSAR-related studies, as does the activity range of six orders of magnitude. However, one weakness might be the lack of structural diversity of most of the compounds.”

Methods

StarDrop’s Auto-Modeller was applied separately to each data set. In each case the data set was split into training, validation and test sets containing 70%, 15% and 15% of the data respectively. This split was performed automatically using the clustering method to ensure the training set represented a good coverage of the structural diversity of the full data set (see the StarDrop Reference Guide for details).

The default descriptors and parameters for descriptor pre-selection were used and all modelling methods provided by the Auto-Modeller were applied.

Details of the parameters and descriptors used are provided in the Auto-Modeller outputs “P-gp inhibitor classification model summary” and “P-gp pEC50 model summary”.

Results

Classification

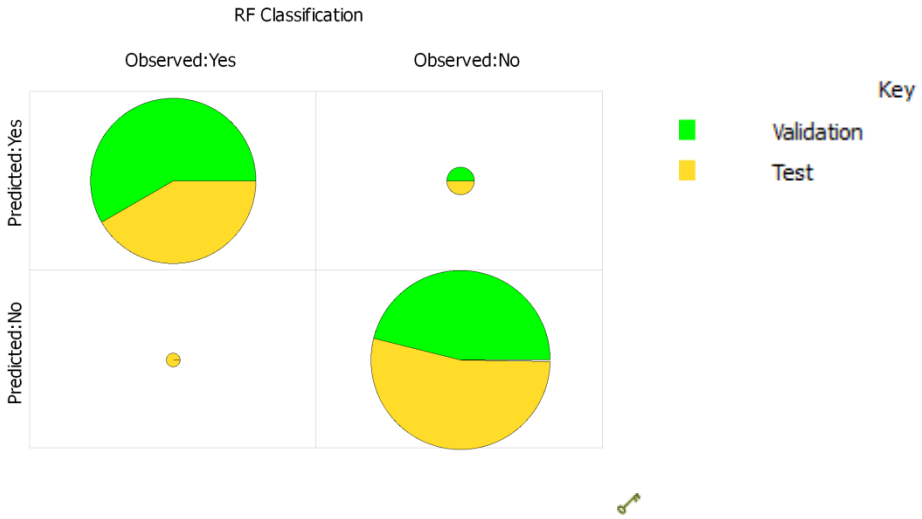
A summary of the results for all of the models built are provided in the Auto-Modeller output “P-gp inhibitor classification model summary.pdf” The best performance on the validation set was achieved by the random forest model. The results are summarised in the table below and full details are provided in the accompanying model output “P-gp_inhibitor_classification_RF.pdf”:

	Training set		Validation set		Test set	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Random forest	1	1	0.93	0.86	0.86	0.71

7221 Cambridge Research Park
Beach Drive, Cambridge
CB25 9TL, UK

Tel: +44 1223 815900
Fax: +44 1223 815907

Email: info@optibrium.com
Website: www.optibrium.com

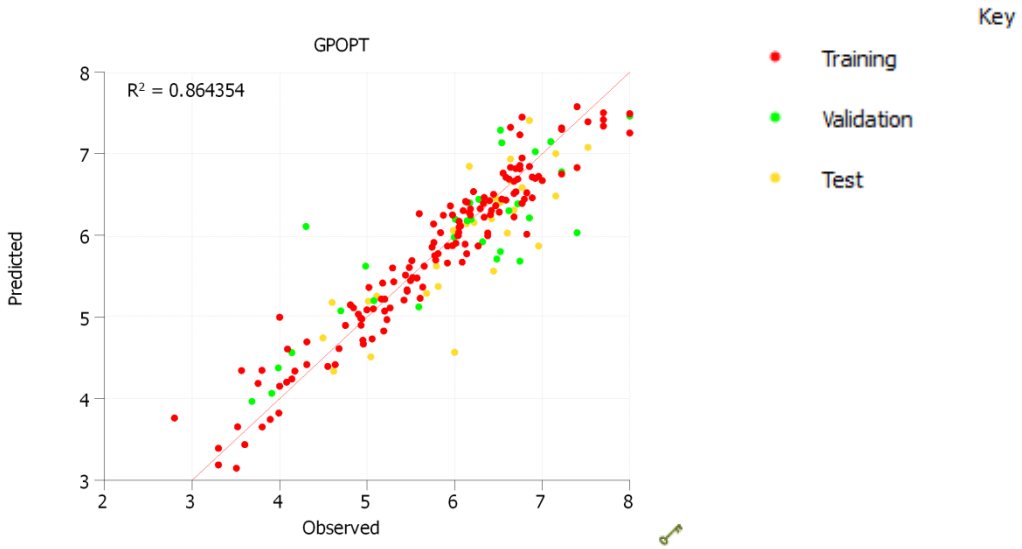


Continuous pEC50

A summary of the results for all of the models built are provided in the Auto-Modeller output "P-gp pEC50 model summary.pdf" The best performance was achieved by the model built with the Gaussian Processes Optimisation method. Details of the model are provided in the accompanying model output "P-gp_pEC50_GPOpt.pdf" and a summary of the results are provided below:

	Training set		Validation set		Test set	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
GP Opt	0.93	0.30	0.70	0.61	0.60	0.50

The R^2 of the test set is slightly less than that for the Validation set. However, this can be understood because the range of observed values for the test set is less than that for the validation set. Note that the RMSE of prediction for these sets are similar.



Using the P-gp Transporter Models

The models can be downloaded for use within StarDrop from the following links:

P-gp_inhibitor_classification.aim

P-gp_pEC50.aim

To use these within StarDrop, download and save these files in a convenient place. Load them into StarDrop using the  button on the **Models** tab. Alternatively, the directory in which the model files have been saved can be added to the paths from which models are automatically loaded when StarDrop starts by selecting the **File->Preference** menu option and adding the directory under **Models** in the **File Locations** tab.

The data sets and detailed outputs from the modelling process may be downloaded in a .zip archive.