

## Caco-2 Model

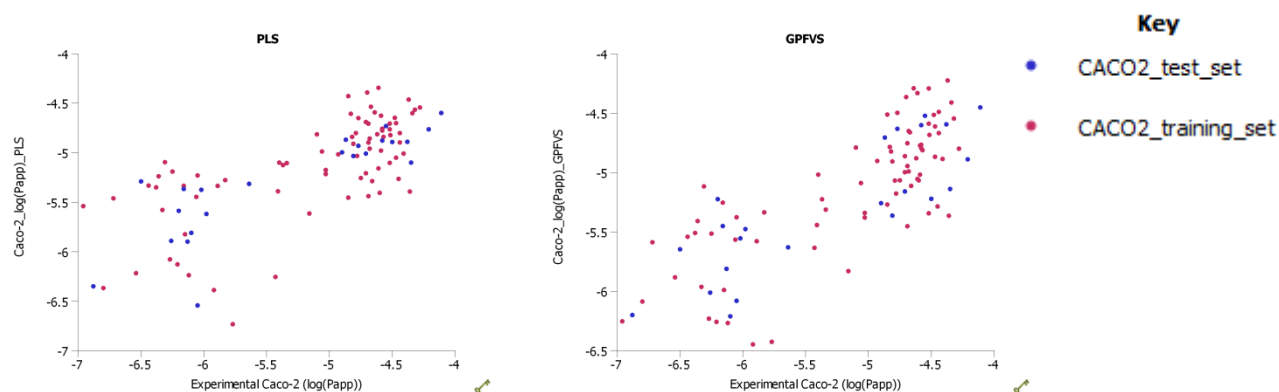
Permeability across monolayers of the Caco-2 line of human epithelial colorectal adenocarcinoma cells is a common *in vitro* model used to assess the potential of compounds to be absorbed across the human intestine. The apparent permeability ( $P_{app}$ ) from the apical to basolateral has been shown to correlate with the *in vivo* fraction absorbed following oral administration of a drug ([Artursson P. and Karlsson J., Biochem. Biophys. Res. Comm.175\(3\), pp. 880–5, 1991](#)). A criterion commonly applied to select compounds with a higher chance of good oral absorption is a  $P_{app}$  value of  $1 \times 10^{-5}$  cm/s or higher.

Nordqvist *et al.* ([QSAR & Comb. Sci. 23\(5\), pp. 303–310, 2004](#)) published experimental Caco-2  $P_{app}$  values for a set of small molecules, including marketed drugs. We have applied the Auto-Modeller to these data to generate models to predict  $\log(P_{app})$ . Two data sets were published with this paper, a training set containing 77 compounds and an independent test set of 23 compounds. For consistency, we have used the same sets to train and validate the models generated.

The training set is smaller than we would ideally like to build a global model of a complex property. To mitigate this, we have used only ten general descriptors to build the models, in an effort to avoid over training. The descriptors used were:  $\log P$ , McGowan's volume, number of hydrogen bond acceptors and donors, flexibility, topological polar surface area, number of aromatic rings, overall charge, negative charge and positive charge. Please see the detailed model output (which can be downloaded below) and the StarDrop Reference Guide for detailed definitions of these descriptors.

All of the modelling techniques available in StarDrop's Auto-Modeller™ were applied and the best models resulted from the PLS and Gaussian Processes (Forward Variable Selection) methods. The results are very comparable to those published in the paper by Nordqvist *et al.* and are summarised in the table below:

Modelling method	Training set			Test set		
	$R^2$	$r^2_{corr}$	RMSE	$R^2$	$r^2_{corr}$	RMSE
PLS	0.47	0.47	0.54	0.66	0.72	0.50
GPFVS	0.60	0.60	0.47	0.66	0.70	0.50




## Using the Caco-2 Models

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

[Caco-2 logPapp\\_PLS.aim](#)

[Caco-2 logPapp\\_GPFVS.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.