

Cytochrome P450 Specificity Module

ACD/Percepta

Model Performance

The Table illustrates performance of the obtained HLM metabolism regioselectivity model on the test set constituting 30% of the data points (atoms) in the initial data set.

For compounds that fall within model Applicability Domain (i.e, obtain acceptable Reliability Index, $RI \geq 0.3$) accuracy and specificity of classification is close higher than 90%. Also, more than a half of the validation set atoms obtain predictions of moderate and high reliability, leading to a substantial improvement in sensitivity and a rise in overall accuracy, which now exceeds 95%.

Table Validation results of the HLM metabolism regioselectivity model. Only metabolism sites predicted with $RI \geq 0.3$ are shown.

Accuracy testing		Calculated probability		Statistical parameters	Overall accuracy
		<0.5	≥ 0.5		
Test set ($RI \geq 0.3$) 2235 atoms (86.6% covered)	Site of no metabolim	1864 (83.4%)	71 (3.2%)	Specificity	96.3%
	Site of metabolim	97 (4.3%)	203 (9.1%)	Sensitivity	67.7%
Test set ($RI \geq 0.5$) 1485 compounds (57.6% covered)	Site of no metabolim	1310 (88.2%)	22 (1.5%)	Specificity	98.3%
	Site of metabolim	34 (2.3%)	119 (8.0%)	Sensitivity	77.8%