

# Cytochrome P450 inhibition (five isoforms)

---

## Dataset profile

The dataset used for the presented CYP inhibition models is a cytochrome panel assay with activity outcomes [[Comprehensive Characterization of Cytochrome P450 Isozyme Selectivity across Chemical Libraries](#); Henrike Veith, Noel Southall, et al.] deposited in the PubChem BioAssay database under identifier [AID1851](#).

The study determined potency values for 17,143 compounds against five CYP isozymes (1A2, 2C9, 2C19, 2D6 and 3A4) using an in vitro bioluminescent assay. The compounds included libraries of US FDA-approved drugs and screening libraries. Among these molecules 8,019 were the compounds from the Molecular Libraries Small Molecule Repository, including compounds chosen for diversity and rule-of-five compliance, synthetic tractability and availability; 6,144 compounds were from biofocused libraries, which included 1,114 FDA-approved drugs; and the rest 2,980 compounds were from combinatorial chemistry libraries, containing privileged structures targeted at G protein-coupled receptors and kinases and containing purified natural products or related structures.

This assay used various human CYP isozymes to measure the dealkylation of various pro-luciferin substrates to luciferin. The luciferin was then measured by luminescence after the addition of a luciferase detection reagent. Pro-luciferin substrate concentration in the assay was equal to its Michaelis constant for its cytochrome P450 isozyme. Inhibitors and some substrates limit the production of luciferin, and decrease measured luminescence. To address potential artifacts due to the assay format, particularly important for panactive compounds, the authors of the assay used a database of potency values determined for the variant of the firefly luciferase used in the assay to remove any compounds that interfered with luciferase detection (only 0.7% were found to be interfering in the compound collection used for the assay).

The activators and compounds marked as inconclusive were removed from the datasets before building the models.

The threshold to differentiate between inhibitors and non-inhibitors is IC<sub>50</sub> of **10  $\mu$ M / L**.

Five models correspond to each of the five cytochrome P450 enzymes measured in the assay.

## Data preprocessing

All chemical structures were cleaned using OCHEM cleaning protocol. The standardization was performed in OCHEM. All salt counter ions were removed and resulting ions were neutralized.

## Descriptors

These models were built using [EState descriptors](#) (electrotopological EState indices) and [ALogPS descriptors](#) according to OCHEM implementation.

## Validation

The models were built and validated using the stratified [bagging](#) technique with 64 bags.

## Statistical parameters

### Prediction accuracy

The basic prediction accuracy parameters according to the bagging validation procedure are:

Cytochrome	Total	Inhibitors	Noninhibitors	Accuracy	Balanced accuracy	MCC	AUC
<b>CYP1A2</b>	13,908 records	6,953	6,955	80.5% $\pm$ 0.3	80.5% $\pm$ 0.3	0.611 $\pm$ 0.007	0.887 $\pm$ 0.01
<b>CYP2C9</b>	13,246 records	4,429	8,817	80.2% $\pm$ 0.3	78.9% $\pm$ 0.4	0.566 $\pm$ 0.008	0.874 $\pm$ 0.01
<b>CYP2C19</b>	13,122 records	5,494	7,634	80.6% $\pm$ 0.3	80.2% $\pm$ 0.3	0.596 $\pm$ 0.007	0.88 $\pm$ 0.01
<b>CYP2D6</b>	14,059 records	2,837	11,222	83% $\pm$ 0.3	76.8% $\pm$ 0.5	0.506 $\pm$ 0.009	0.839 $\pm$ 0.01
<b>CYP3A4</b>	15,334 records	5,819	9,515	80.6% $\pm$ 0.3	80.2% $\pm$ 0.3	0.596 $\pm$ 0.007	0.88 $\pm$ 0.01

### Applicability domain

The prediction accuracy is estimated using PROB-STD distance to model and sliding window based accuracy averaging.